

SEQ ID NO # 2

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:19:12 ; Search time 51.87 seconds

(without alignments)
42,828 Million cell updates/sec

Title: US-09-582-296-2

Sequence: 1 APAYVMDAESEFALAHGGL 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_032802:*
1: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	101	100.0	20	AAV27011	N-terminal fragmen
2	69	68.3	18	AAV27010	N-terminal fragmen
3	51.5	51.0	22	AAV27011	C glutamicum prote
4	48	47.5	22	AAU01898	Mycobacterium tube
5	45.5	45.0	22	ABB65488	Drosophila melanog
6	45.5	45.0	22	ABB65489	Drosophila melanog
7	45	44.6	14	AAU49925	Consensus haemaggl
8	44.5	44.1	474	AAU3631	Pseudomonas aerugi
9	43	42.6	130	AAV27011	Rice glutaredoxin
10	43	42.6	317	AAV74126	Human prostate tum
11	43	42.6	330	AAW94066	Human DnaJ-like pr

12	43	42.6	524	20	AAW94753	Mutant H protein a
13	43	42.6	617	14	AAK34541	H protein of atten
14	43	42.6	617	14	AAK42387	Moratan haemagglut
15	43	42.6	617	14	AAK42388	Consensus haemaggl
16	43	42.6	617	14	AAK42389	San Diego haemaggl
17	43	42.6	617	14	AAK42390	Chicago 1 haemaggl
18	43	42.6	617	14	AAK42391	Chicago 2 haemaggl
19	43	42.6	617	14	AAK42392	McI haemagglutinin
20	43	42.6	617	14	AAK42393	McI haemagglutinin
21	43	42.6	617	20	AAW94752	Mutant measles vir
22	42.5	42.1	385	22	AAAG90134	C glutamicum prote
23	42	41.6	106	22	ABG01349	Novel human diagno
24	42	41.6	148	22	ABG01349	Novel human diagno
25	42	41.6	180	22	AAAG6383	Human partial oita
26	42	41.6	216	22	AAAG72004	Human olfactory re
27	42	41.6	216	22	AAAG72923	Human olfactory re
28	42	41.6	216	22	AAAG72924	Human olfactory re
29	42	41.6	216	22	AAAG72926	Human olfactory re
30	42	41.6	228	22	ABG01309	Novel human diagno
31	42	41.6	239	22	ABG01350	Novel human diagno
32	42	41.6	301	22	AAAG72927	Human olfactory re
33	42	41.6	305	22	AAAG6367	Human partial oita
34	42	41.6	305	22	AAAG6389	Human partial oita
35	42	41.6	317	22	AAAG71408	Human olfactory re
36	42	41.6	317	22	AAAG72330	Human olfactory re
37	42	41.6	317	22	AAAG72925	Human olfactory re
38	42	41.6	317	22	AAAG72927	Human olfactory re
39	42	41.6	584	22	ABG68361	Drosophila melanog
40	42	41.6	608	22	AAAG1551	C glutamicum prote
41	41	40.6	116	21	AAV50927	Human fetal brain
42	41	40.6	131	21	AAV57124	Human prostate can
43	41	40.6	143	21	AAV57124	Human ORF222
44	41	40.6	185	21	AAV57124	Arabidopsis thalia
45	41	40.6	185	22	AAAG1571	C glutamicum prote

ALIGNMENTS

RESULT 1	
AAV27011	AAV27011 standard; peptide: 20 AA.
ID	AAV27011; (first entry)
AC	AAV27011; (first entry)
XX	24-SEP-1999 (first entry)
XX	N-terminal fragment of anti-freeze protein from Lichen.
XX	Antifreeze protein; AFP; Lichen; frozen food product; pasteurisation;
XX	sterilisation; freezing; frozen confectionery product; N-terminal.
XX	Umbilicaria antarctica.
XX	WO9937673-A2.
XX	29-JUL-1999.
XX	23-DEC-1998; 98WO-EP08554.
XX	22-JAN-1998; 98GB-0001420.
XX	(UNL) UNILEVER NV.
XX	(UNL) UNILEVER PLC.
XX	Byass LJ, Sidebottom CM, Smallwood MF;
XX	WPI, 1999-444595/37.
XX	New isolated antifreeze protein obtained from Lichen, used for the
XX	preparation of food products, particularly frozen confectionery
XX	products

PS Claim 2; Page 17; 20pp; English.

CC The invention describes a novel antifreeze protein (AFP) obtained from
 CC Lichen that comprises an apparent mol. wt. of 20-28kD and has an
 CC N-terminal amino acid sequence which shows at least 80% overlap with
 CC sequence shown in AAY27010; and modified versions and isoforms of
 CC this protein. The AFP can conveniently be used in food products,
 CC preferably in food products which are frozen or intended to be frozen.
 CC Especially preferred is the use of AFPs in products which are heated e.g.
 CC by pasteurisation or sterilisation prior to freezing and in frozen
 CC confectionery products. Using the AFP ingredient, mixes can be frozen
 CC under quiescent conditions, e.g. in a shop or home freezer without the
 CC formation of unacceptable ice crystal shapes and hence with a texture
 CC different to products normally obtained via quiescent freezing. The
 CC present sequence represents a N-terminal fragment of an 24 kDa AFP.

XX Sequence 20 AA;

Query Match 100.0%; Score 101; DB 20; Length 20;

Best Local Similarity 100.0%; Pred. No. 1,1e-10;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 APAYVMGDAESFGAIAHGCL 20
 |||
 Db 1 apayvmgdaesfgaiahgcl 20

RESULT 2

AAY27010 AAY27010 standard; peptide; 18 AA.

XX AAY27010;

XX 24-SEP-1999 (first entry)

XX N-terminal fragment of anti-freeze protein from Lichen.

XX Antifreeze protein; AFP; Lichen; frozen food product; pasteurisation;
 KW sterilisation; freezing; frozen confectionery product; N-terminal.

XX Umbilicaria antarctica.

XX MO9937673-A2.

XX 29-JUL-1999.

XX 23-DEC-1998; 98WO-EP08554.

XX 22-JAN-1998; 98GB-0001420.

XX (UNIL) UNILEVER NV.

XX (UNIL) UNILEVER PLC.

XX Byass LJ, Sidebottom CM, Smallwood MF;

XX WPI; 1999-444595/37.

XX New isolated antifreeze protein obtained from Lichen, used for the
 PT preparation of food products, particularly frozen confectionery
 PT products

XX Claim 1; Page 17; 20pp; English.

CC The invention describes a novel antifreeze protein (AFP) obtained from
 CC Lichen that comprises an apparent mol. wt. of 20-28kD and has an
 CC N-terminal amino acid sequence which shows at least 80% overlap with
 CC the present sequence; and modified versions and isoforms of this
 CC protein. The AFP can conveniently be used in food products, preferably
 CC in food products which are frozen or intended to be frozen. Especially
 CC preferred is the use of AFPs in products which are heated e.g. by
 CC pasteurisation or sterilisation prior to freezing and in frozen
 CC confectionery products. Using the AFP ingredient, mixes can be frozen

CC under quiescent conditions, e.g. in a shop or home freezer without the
 CC formation of unacceptable ice crystal shapes and hence with a texture
 CC different to products normally obtained via quiescent freezing. The
 CC present sequence represents a N-terminal fragment of an AFP.

XX Sequence 18 AA;

Query Match 68.3%; Score 69; DB 20; Length 18;

Best Local Similarity 80.0%; Pred. No. 3.6e-05;
 Matches 16; Conservative 1; Mismatches 1; Indels 2; Gaps 1;

QY 1 APAYVMGDAESFGAIAHGCL 20
 |||
 Db 1 apayvm--daesfgaiahgcl 18

RESULT 3

AAG92771 AAG92771 standard; Protein; 751 AA.

XX AAG92771;

XX 26-SEP-2001 (first entry)

XX C glutamicum protein fragment SEQ ID NO: 6525.

XX Coryneform bacterium; amino acid synthetase; vitamin; saccharide;
 KW organic acid synthesis.

XX Corynebacterium glutamicum.

XX EP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX N-PSDB; AAH67990.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 17; SEQ ID NO: 6525; 246pp + Sequence Listing; English.

CC The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX Sequence 751 AA;

Query March	44.1%	Score 44.5	DB 22	Length 474
Best Local Similarity	50.0%	Pred. No. 28		
Matches 10; Conservative	3;	Mismatches	2;	Indels 5; Gaps 1
Qy	4	VWVGDAE-----SEGAIAHG	18	
	:	: :		
Db	130	vvcgdsmtstchgafgatalng	149	

RESULT	9
AAB29756	
ID	AAB29756 standard; Protein; 130 AA.

DT 28-FEB-2001 (first entry)

Rice glutaredoxin homologue 2 (GRL2), SEQ ID NO:16.

KM Rice; GR22; glutaredoxin homologue 2; MNL interactor; SAR; bZIP protein.
KM systemic acquired resistance; yeast two-hybrid system; transgenic plant.
KM pathogen resistance; virus; bacterium; nematode; fungus; insect.

OS *Oryza sativa*

PN WO2000063417-A2

PD 26-OCT-2000.

PF 06-APR-2000; 2000WO-US09060.

PR 19-APR-1999; 99US-0294539.

PA (REGC) UNIV CALIFORNIA.

DR WPI; 2001-090867/10.

XX

PT rice proline rich NPR1 interactor proteins, polypeptides interacting

2
1
4
3
2
2
2
1
2
2
2

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

nucleotides encoding the protein in the induction of autoimmunity.

resistance response that can be induced during a local infection by an avirulent pathogen. A rice cDNA library was screened using the wheat two

PNI (proline rich NPRO1 interactor: AAB29749) and MN1 (AAB29758) were

CC was then rescreened using rice PNI cDNA or rice MN1 cDNA as bait. Rice

CC protein (AAB29752), PREG-like protein (AAB29753), and a protein with no

CC protein, and rice GRL1 and GRL2 (glutaredoxin homologues 1 and 2;

microtubule-associated protein; AAB29757) were found to interact with

plants comprising an expression cassette encoding a protein of the invention and a method for obtaining the same.

by introducing the expression cassette into the plant, and selecting the

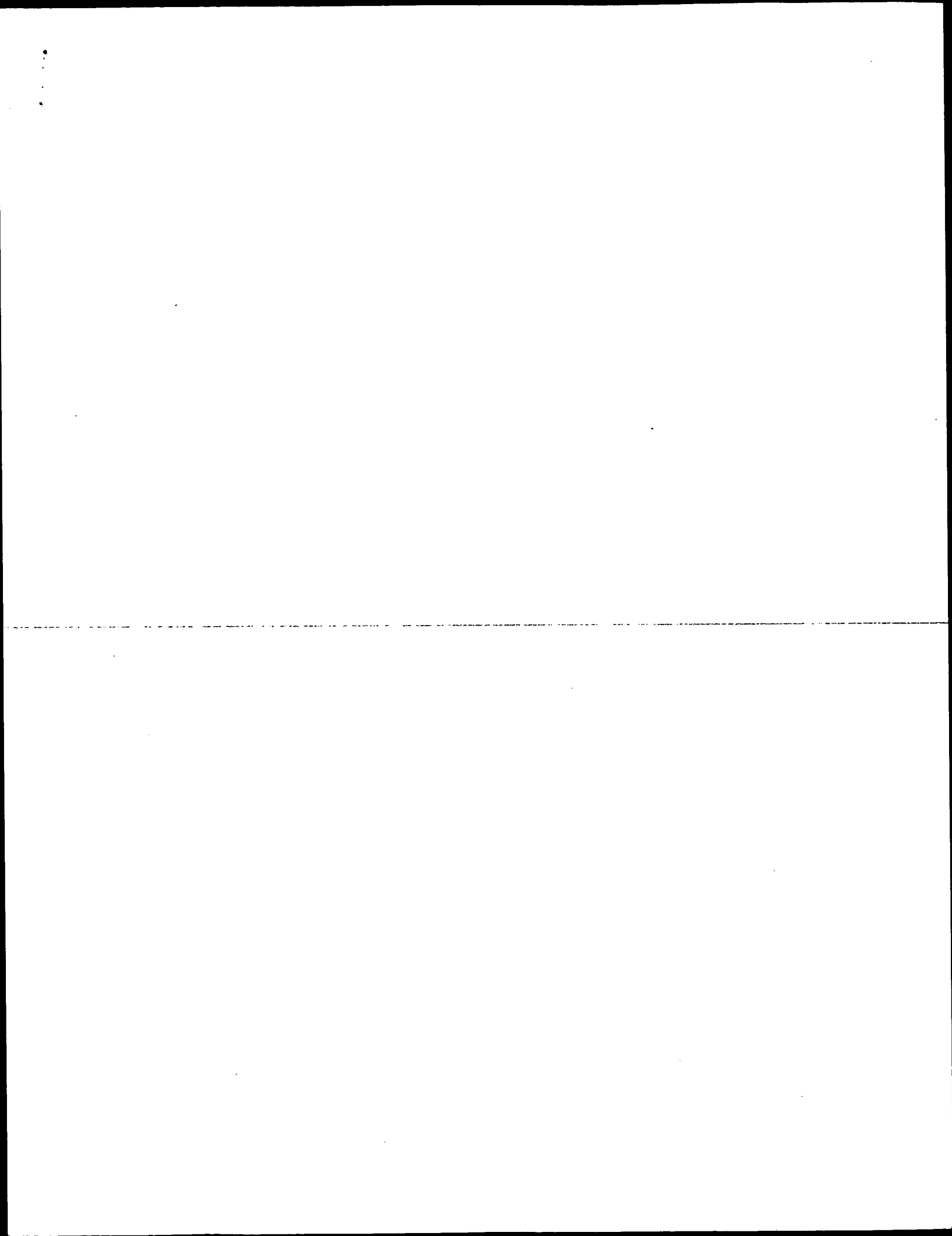
CC proteins are useful for enhancing plant resistance to pathogens such as
CC viruses bacteria nematodes fungi or insects. The present disclosure

cc represents ICE GRLZ.
xx

SQ Sequence 130 AA;

Mon Jul 8 09:33:16 2002

us-09-582-296-2.rag



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:17:25 ; Search time 41.68 Seconds
(without alignments)
74.710 Million cell updates/sec

Title: US-09-582-296-1
Perfect score: 98
Sequence: 1 APAMDAESFGAIGHGL 18

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 segs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_protist:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriophage:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	57.1	145	2	09K4B2
2	50	51.0	963	13	042110
3	49	50.0	461	11	09DC17
4	49	50.0	461	11	09D728
5	49	50.0	461	11	099KR8
6	49	50.0	467	4	09BTV2
7	49	50.0	467	4	09UJM5
8	48	49.0	116	11	09D242
9	48	49.0	452	11	09DD22
10	48	49.0	572	11	099LJ1
11	48	49.0	572	11	042728
12	47	48.0	583	16	09PHM7
13	46	46.9	341	4	09P133
14	46	46.9	400	4	09NT44
15	46	46.9	628	4	09HB63
16	46	46.9	628	4	09B2P1

Seq ID no 1

17	45	45.9	363	16	091505	091595 pseudomonas
18	45	45.9	384	16	069579	069579 mycobacteri
19	45	45.9	449	16	09WYE2	09WYE2 thermotoga
20	44	44.9	281	16	098L72	098L72 rhizobium 1
21	44	44.9	421	16	092502	092502 rhizobium m
22	44	44.9	517	16	09CJ30	09CJ30 lactococcus
23	43.5	44.4	258	16	007601	007601 bacillus su
24	43.5	44.4	791	2	09L115	09L115 streptomyce
25	43	43.9	212	10	09ARX5	09ARX5 oryza sativ
26	43	43.9	246	16	0987Y7	0987Y7 rhizobium 1
27	43	43.9	250	4	09B3H1	09B3H1 homo sapien
28	43	43.9	250	4	096159	096159 homo sapien
29	43	43.9	263	10	09FR70	09FR70 oryza sativ
30	43	43.9	425	16	098L43	098L43 rhizobium 1
31	42.5	43.4	769	16	09A282	09A282 caulobacter
32	42.5	43.4	1222	5	09VF52	09VF52 dirosophila
33	42.5	43.4	334	17	029369	029369 archaeoglob
34	42	42.9	483	16	09RTM0	09RTM0 delinococcus
35	42	42.9	492	2	051065	051065 neisseria g
36	42	42.9	541	3	09P6T4	09P6T4 neurospora
37	42	42.9	547	10	09MA11	09MA11 arabidopsis
38	42	42.9	617	13	090ZF6	090ZF6 oryzae lat
39	42	42.9	700	13	091372	091372 xenopus lae
40	42	42.9	899	5	09VND0	09VND0 dirosophila
41	42	42.9	1486	16	09CG24	09CG24 lactococcus
42	41.5	42.3	233	16	098A01	098A01 rhizobium 1
43	41.5	42.3	474	12	089542	089542 bovine herp
44	41.5	42.3	518	2	09RKZ7	09RKZ7 streptomyce
45	41.5	42.3				

ALIGNMENTS

RESULT 1

ID 09K4B2 PRELIMINARY; PRT; 145 AA.

AC 09K4B2;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE HYPOTHETICAL 15.4 KDA PROTEIN.

GN SC7E4.04C.

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.

OX NCBI_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA Seeger K.J., Harris D.;

RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajadream M.A.;

RL Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-A3(2);

RX MEDLINE=97000351; PubMed=8843436;

RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,

RT Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for

RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

DR EMBL; AL359214; CAB94592.1; -

DR InterPro: IPR000644; CBS.

DR Pfam: PF00571; CBS; 2.

DR SMART; SM00116; CBS; 2.

KW Hypothetical protein.

SO SEQUENCE 145 AA; 15353 MW; FAE58BBA18E358B5 CRC64;

```

Query Match          57.1%; Score 56; DB 2; Length 145;
Best Local Similarity 58.8%; Pred. NO. 0.2;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 APAMDAESFGAIAHG 17
    ||| 1:|||||
DB 90 APTWVEAGAGAMAHG 106

RESULT 2
Q42110 PRELIMINARY; PRT; 963 AA.
AC 042110;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ORF2.
OS Platemys spixii.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Testudines; Pleurodira; Chelidae; Platemys.
OX NCBI_TaxID=45338;
RN [1]
RC SEQUENCE FROM N.A.
RP TRANSDON-RETROTRANSDON CRI-LIKE LINE (PSCR1);
RX MEDLINE=98066352; PubMed=9402732;
RA Kajikawa M., Ohshima K., Okada N.;
RT "Determination of the entire sequence of turtle CRI: the first open
RT reading frame of the turtle CRI element encodes a protein with a novel
RT zinc finger motif.";
RL MOL. BIOL. EVOL. 14:1206-1217(1997).
DR EMBL: AB005891; BAA8337.1; -.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00078; rvt, 1.
KM RNA-directed DNA polymerase.
SQ SEQUENCE 963 AA; 109601 MW; 2842A81F9B6487D CRC64;

Query Match          51.0%; Score 50; DB 13; Length 963;
Best Local Similarity 50.0%; Pred. NO. 17;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 PAMMDAESFGAIAH 15
    ||| 1:|
DB 333 PAMLNKXKGLKH 346

RESULT 3
Q9DCL7 PRELIMINARY; PRT; 461 AA.
AC 09DCL7;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 0610025011RIK.
GN 0610025011RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=KIDNEY;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischer W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schmitt L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

```

```

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK002675; BAB22277.1; -.
DR MGD: MGI:1914098; 0610025011RIK.
DR InterPro: IPR000933; Alpha.L.fucos.
DR Pfam: PF01120; Alpha.L.fucos. 1.
DR PRINTS: PR00741; GLHTRLASE29.
DR PROSITE: PS00385; ALPHA_L_FUCOSIDASE; UNKNOWN_1.
SQ SEQUENCE 461 AA; 53631 MW; 1E233114026A3C1F CRC64;

Query Match          50.0%; Score 49; DB 11; Length 461;
Best Local Similarity 47.1%; Pred. NO. 10;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 2 PAMMDAESFGAIAHGL 18
    ||| 1:|
DB 41 PAMFQAKFGFIHWGV 57

RESULT 4
Q9D728 PRELIMINARY; PRT; 461 AA.
AC 09D728;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 0610025011RIK.
GN 0610025011RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=STOMACH;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischer W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schmitt L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK008653; BAB25809.1; -.
DR MGD: MGI:1914098; 0610025011RIK.
DR InterPro: IPR000933; Alpha.L.fucos.
DR Pfam: PF01120; Alpha.L.fucos. 1.
DR PRINTS: PR00741; GLHTRLASE29.
DR PROSITE: PS00385; ALPHA_L_FUCOSIDASE; UNKNOWN_1.
SQ SEQUENCE 461 AA; 53530 MW; 25754D6F0F24D09B CRC64;

```


RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 DR EMBL: AK020610; BAB32146.1; -;
 DR MGD: MGI:1925799; 9530055J05R1K.
 SQ SEQUENCE 116 AA; 13558 MW; F4E7E9AE6306A2A CRC64;

Query Match 49.0%; Score 48; DB 11; Length 116;
 Best Local Similarity 41.2%; Pred. No. 3.1;
 Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 PAMMAESFGAIAHGCL 18
 Db 36 PSMFEAKFGVFMHGV 52

RESULT 9
 ID 09DD22 PRELIMINARY; PRT; 452 AA.
 AC 09DD22;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE 0610006A03RIK PROTEIN.
 GN 0610006A03RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCB1_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=KIDNEY;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa Y., Nishi K., Kiyosawa H., Kondo S., Yamanka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Felschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
 RA Schiml L.M., Stahli F., Suzuki R., Tomita M., Wagner L., Mashio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlaya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
 RA Wyszaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection";
 RL Nature 409:685-690(2001).
 DR EMBL: AK002230; BAB21949.1; -;
 DR MGD: MGI:1918915; 0610006A03RIK.
 DR InterPro: IPR000933; Alpha_L_Fucos.
 DR Pfam: PF01120; Alpha_L_Fucos; 1.
 DR PROSITE: PR00741; GLHYDRLASE29.
 DR PROSITE: PS00385; ALPHA_L_FUCOSIDASE; 1.
 SQ SEQUENCE 452 AA; 52314 MW; 4BB919AA4A152B2C CRC64;

Query Match 49.0%; Score 48; DB 11; Length 452;
 Best Local Similarity 41.2%; Pred. No. 15;
 Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Db 36 PSMFEAKFGVFMHGV 52
 RESULT 10
 ID 099LJ1 PRELIMINARY; PRT; 452 AA.
 AC 099LJ1;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE RIKEN CDNA 0610006A03 GENE.
 GN 0610006A03RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCB1_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD. CROSS
 RC TISSUE:;
 RA Strausberg R.,
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC003235; AAH03235.1; -;
 DR MGD: MGI:1918915; 0610006A03RIK.
 DR InterPro: IPR000933; Alpha_L_Fucos.
 DR Pfam: PF01120; Alpha_L_Fucos; 1.
 DR PROSITE: PR00741; GLHYDRLASE29.
 DR PROSITE: PS00385; ALPHA_L_FUCOSIDASE; 1.
 SQ SEQUENCE 452 AA; 52280 MW; 41B319A040152B2C CRC64;

Query Match 49.0%; Score 48; DB 11; Length 452;
 Best Local Similarity 41.2%; Pred. No. 15;
 Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 PAMMAESFGAIAHGCL 18
 Db 36 PSMFEAKFGVFMHGV 52
 RESULT 11
 ID 042728 PRELIMINARY; PRT; 572 AA.
 AC 042728;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CEPHALOSPORIN ESTERASE.
 OS Rhodospirillum rubrum (Yeast) (Rhodotorula gracilis).
 OC Eukaryota; Fungi; Basidiomycota; Urediniomycetes;
 OC Microbotryomycetidae; Heterogastriales; Sporidiobolaceae;
 OC Rhodospirillum.
 OX NCB1_Taxid=5286;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=98069473; PubMed=9406399;
 RX Polittine M., Tonzi S.M., Burnett W.V., Romanick G., Usher J.J.,
 RT "Purification and characterization of a cephalosporin esterase from
 RT Rhodospirillum rubrum";
 RL Appl. Environ. Microbiol. 63:4807-4811(1997).
 CC -1- SIMILARITY: BELONGS TO THE TYPE-B CARBOXYLESTERASE/LIPASE FAMILY.
 DR EMBL: AF025410; AAB93483.1; -;
 DR HSP: P37967; 10E3.
 DR InterPro: IPR002018; Carboxylesterase_B.
 DR InterPro: IPR000379; Est_Lip_thioest_actsite.
 DR Pfam: PF00135; Coesterase; 1.
 DR PROSITE: PS00122; CARBOXYLESTERASE_B_1; 1.
 DR KW Hydroxylase; Porin.
 SQ SEQUENCE 572 AA; 61315 MW; CC104DOC7682997E CRC64;

Query Match 49.0%; Score 48; DB 3; Length 572;

Best Local Similarity 52.9%; Pred. No. 19;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAMDAESFGAIAHGCL 18
DB 442 PTYTAEGFSSAHKGL 458

RESULT 12

Q9PHM7 PRELIMINARY; PRT; 583 AA.
AC Q9PHM7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ASPARTYL-TRNA SYNTHETASE (EC 6.1.1.12).
GN ASPS OR C00640.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIC 11168;
RX MEDLINE=20150912; PubMed=1068204;
RA Parish J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holroyd S.,
RA Jagels K., Kallish A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajadream M.A., Rutherford K.M., Van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -1- CATALYTIC ACTIVITY: ATP + L-AMINO ACID + TRNA(AMINO ACID) = AMP +
CC PYROPHOSPHATE + L-AMINOACYL-TRNA(AMINO ACID).
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE (OR L-ASPARAGINE) +
CC TRNA(ASN) = AMP + PYROPHOSPHATE + L-ASPARTYL-TRNA(ASP) (OR L-
CC ASPARAGINYL-TRNA(ASN)).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -1- SIMILARITY: BELONGS TO ASPARTYL-TRNA SYNTHETASE FAMILY.
DR EMBL; AL139075; CAB75276.1; -.
DR HSSP; P21889; IEOR.
DR InterPro; IPR002106; AA_TRNA_Ligase_II.
DR InterPro; IPR002312; trna-synt_ase.
DR InterPro; IPR002313; trna-synt_lys_2.
DR InterPro; IPR004363; trna_antl.
DR Pfam; PF01336; trna_antl; 1.
DR PRINTS; PR01042; TRNASYNTHASP.
DR PRINTS; PR00982; TRNASYNTHLVS.
DR PROSITE; PS00179; AA_TRNA_Ligase_II_1; 1.
DR PROSITE; PS00339; AA_TRNA_Ligase_II_2; UNKNOWN_1.
DR ATP-binding; Aminoacyl-TRNA synthetase; Complete proteome; Ligase;
KW Protein biosynthesis.
SQ SEQUENCE 583 AA; 66166 MW; A3723A2E100222E2 CRC64;

Query Match 48.0%; Score 47; DB 16; Length 583;
Best Local Similarity 64.38; Pred. No. 29;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 MDAESFGAIAHGCL 18
DB 513 LDALSGAPPHGCI 526

RESULT 13

Q9P133 PRELIMINARY; PRT; 341 AA.
AC Q9P133;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PRO3091 (NEIRIN 4).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Zhang C., Yu Y., Zhang S., Wei H., Zhang Y., Zhou G., Bi J., Liu M.,
RA He F.;
RT "Functional prediction of the coding sequences of 79 new genes deduced
RT by analysis of cDNA clones from human fetal liver."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hiki J., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF119916; AAF69670.1; -.
DR EMBL; AK024691; BAB14964.1; -.
DR EMBL; BC013591; AAH13591.1; -.
DR HSSP; P02468; ITLE.
DR InterPro; IPR000561; EGF-like_1.
DR InterPro; IPR002049; laminin_EGF.
DR InterPro; IPR001134; Nectin_C.
DR Pfam; PF00053; laminin_EGF; 2.
DR Pfam; PF01759; NTR; 1.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SM00180; EGF_Iam; 2.
DR SMART; SM00001; EGF_Like; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_2.
SQ SEQUENCE 341 AA; 37907 MW; 13E60C3ABBFC87 CRC64;

Query Match 46.9%; Score 46; DB 4; Length 341;
Best Local Similarity 50.08; Pred. No. 23;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 WMDAESFGAIAHGCL 17
DB 204 WMDAEGFSAIAHGCL 217

RESULT 14

Q9NT44 PRELIMINARY; PRT; 400 AA.
AC Q9NT44;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOPHYSICAL 44.6 KDA PROTEIN (FRAGMENT).
GN DKFZP434O1519.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ottenmaier B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.,
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL137540; CAB70800.1; -.
DR HSSP; P02468; ITLE.
DR InterPro; IPR000561; EGF-like.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: July 8, 2002, 08:18:20 ; Search time 13.48 Seconds

(without alignments)
51.703 Million cell updates/sec

Title: US-09-582-296-1
Perfect score: 98
Sequence: 1 APAMDAEFGAIHGGI 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	50	51.0	461 1	P40066 homo sapien
2	49	50.0	407 1	P43713 caenorhabdi
3	46	46.9	578 1	D5BD_PSESP
4	45	45.9	462 1	FUCO_RAT
5	44.5	45.4	227 1	DUB8_MOUSE
6	44	44.9	465 1	FUCO_CANPA
7	44	44.9	545 1	SYN_CABEL
8	43	43.9	326 1	DDB6_HUMAN
9	42.5	43.4	807 1	YNF_PSEAE
10	42	42.9	461 1	FUCO_DICDI
11	42	42.9	591 1	SYD_PSEAE
12	41	41.8	448 1	SYN_STRPY
13	40	40.8	97 1	VGIC_BPPZA
14	40	40.8	160 1	BCHF_RHOSH
15	40	40.8	171 1	BCHF_RHOSH
16	40	40.8	173 1	LMIP_BOVIN
17	40	40.8	185 1	ISFZ_HAEIN
18	40	40.8	256 1	YAFV_ECOLI
19	40	40.8	317 1	OZEL_HUMAN
20	40	40.8	447 1	SYN_STRPN
21	40	40.8	587 1	REC_N_MYCU
22	40	40.8	589 1	SYD_MYCLE
23	40	40.8	1515 1	GLTB_AZOB
24	39.5	40.3	209 1	ENMD_ECOLI
25	39.5	40.3	294 1	DAPA_CAUCR
26	39.5	40.3	1335 1	PSBB_MARO
27	39.5	40.3	1335 1	PSBB_MARO
28	39	39.8	146 1	IKB_TRIFL
29	39	39.8	242 1	DJB6_MOUSE
30	39	39.8	342 1	EFTS_LACIA
31	39	39.8	345 1	EFTS_STRPN
32	39	39.8	346 1	Y193_HUMAN
33	39	39.8	389 1	VMSA_HPBVA

ALIGNMENTS

RESULT 1	ID	FUCO_HUMAN	STANDARD:	PRT:	461 AA.
AC	P04066	Q14334	Q14335		
DT	01-NOV-1996	(Rel. 03, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Tissue alpha-L-fucosidase precursor (EC 3.2.1.51) (Alpha-L-fucosidase I) (Alpha-L-fucoside fucosyltransferase).				
CN	FUCAL				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=90026416				
RA	Ochiodoro T., Beckmann K.R., Morris C.P., Hopwood J.J.;				
RT	"Human alpha-L-fucosidase: complete coding sequence from cDNA clones."				
RL	Biochem. Biophys. Res. Commun. 164:439-445(1989).				
RN	[2]				
RP	SEQUENCE OF 70-421 FROM N.A.				
RX	MEDLINE=88151864				
RA	O'Brien J.S., Williams P.J., Fukushima H., de Wet J.R., Darby J.K.,				
RT	DiCioccio R., Fowler M.L., Shows T.B.;				
RL	"Molecular biology of the alpha-L-fucosidase gene and fucosidosis."				
RN	[3]				
RP	SEQUENCE OF 62-462 FROM N.A.				
RX	MEDLINE=85140268				
RA	Fukushima H., de Wet J.R., O'Brien J.S.;				
RT	"Molecular cloning of a cDNA for human alpha-L-fucosidase."				
RL	Proc. Natl. Acad. Sci. U.S.A. 82:1262-1265(1985).				
RN	[4]				
RP	SEQUENCE OF 70-422 FROM N.A.				
RX	MEDLINE=85076174				
RA	de Wet J.R., Fukushima H., Dewji N.N., Wilcox E., O'Brien J.S.,				
RT	Helinski D.R.;				
RL	"Chromogenic immunodetection of human serum albumin and alpha-L-fucosidase clones in a human hepatoma cDNA expression library."				
RN	[5]				
RP	STRUCTURE OF CARBOHYDRATES.				
RA	Beem E.P., Lismann J.J.W., van Steijn G.J., van der Wal C.J.,				
RT	Trippelvitiz L.A.W., Overdijk B., van Halbeek H., Mutsaers J.H.G.M.,				
RL	Vliegenterhart J.F.G.;				
RN	[6]				
RP	STRUCTURAL ANALYSIS OF THE CARBOHYDRATE MOETIES OF ALPHA-L-FUCOSIDASE FROM HUMAN LIVER."				
RA	Glycoconj. J. 4:33-42(1987).				
RN	[6]				
RP	VARIANT ARG-281.				
RX	MEDLINE=94002224				
RA	Yang M., Allen H., DiCioccio R.A.;				
RT	"Pedigree analysis of alpha-L-fucosidase gene mutations in a fucosidosis family."				

34	39	39.8	397 1	MPA2_AMBAR	P27762 ambrosia ar
35	39	39.8	481 1	GLNA_HELPY	P94845 heliobacte
36	39	39.8	1051 1	UBA1_WHEAT	P20973 triticum ae
37	39	39.8	1051 1	UBA2_WHEAT	P31251 triticum ae
38	39	39.8	2194 1	GLSN_MEDSA	003460 medicago sa
39	38.5	39.3	430 1	SYN_BACHD	09K078 bacillus ba
40	38.5	39.3	442 1	STD_ABRPE	03Y907 aeropyrum p
41	38.5	39.3	527 1	PUR1_MYCTU	006626 mycobacteri
42	38.5	39.3	556 1	PUR1_MYCTU	050028 mycobacteri
43	38.5	39.3	636 1	FTHS_SPTOL	P28723 spiniacia ol
44	38	38.8	278 1	YKF5_YEAST	P35731 saccharomyc
45	38	38.8	389 1	VMSA_HPBVA	P03138 hepatitis b

RL Blochim Biophys. Acta 1182:245-249(1993).
[17]
RP VARIANT FUCOSIDOSIS ASP-60. AND VARIANT ARG-281.
RX MEDLINE=93278392; PubMed=8504303;
RA Seo H.-C., Willems P.J., Kreiz K.A., Martin B.M., O'Brien J.S.;
RT "Fucosidosis: four new mutations and a new polymorphism.";
RL Hum. Mol. Genet. 2:423-429(1993).
[18]
RP VARIANT FUCOSIDOSIS LEU-63.
RX MEDLINE=95179129; PubMed=7874128;
RA Seo H.-C., Yang M., Tonlorenzi R., Willems P.J., Kim A.H.,
RT Filocomo M., Gatti R., DiGiocco R.A., O'Brien J.S.;
RT "A missense mutation (S63I) in alpha-L-fucosidase is responsible for
fucosidosis in an Italian patient."
RL Hum. Mol. Genet. 3:2065-2066(1994).
[19]
RN VARIANT ARG-281.
RM MEDLINE=95114998; PubMed=7815431;
RX Cregg H., Winchester B., Seo H.C., O'Brien J., Swallow D.;
RT "Molecular basis of the common electrophoretic polymorphism (Fu1/Fu2)
in human alpha-L-fucosidase."
RL J. Med. Genet. 31:659-660(1994).
CC -1 FUNCTION: ALPHA-L-FUCOSIDASE IS RESPONSIBLE FOR HYDROLYZING THE
ALPHA-1,6-LINKED FUCOSE JOINED TO THE REDUCING-END
N-ACETYLGLUCOSAMINE OF THE CARBOHYDRATE MOETIES OF GLYCOPROTEINS.
CC -1 CATALYTIC ACTIVITY: An alpha-L-fucose + H(2)O = an alcohol + L-
fucose.
CC -1 SUBUNIT: HOMOTETRAMER.
CC -1 SUBCELLULAR LOCATION: Lysosomal.
CC -1 POLYMORPHISM: THE VARIATION IN POSITION 281 IS RESPONSIBLE FOR
THE FU1/FU2 POLYMORPHISM.
CC -1 DISEASE: A DEFICIENCY OF THIS ACID HYDROLASE IN HUMANS AND DOGS
RESULTS IN THE LYSOSOMAL STORAGE DISEASE FUCOSIDOSIS.
CC -1 SIMILARITY: BELONGS TO FAMILY 29 OF GLYCOSYL HYDROLASES.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss institute of Bioinformatics and the EMBL outstation at
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/isb-sib.ch>).
or send an email to license@isb-sib.ch.
CC -----
DR EMBL; M29877; AAA35519.1; .
DR EMBL; X01390; CA825646.1; .
DR EMBL; M10355; AAA52482.1; .
DR EMBL; M80815; AAA52481.1; .
DR EMBL; M80810; AAA52481.1; JOINED.
DR EMBL; M80809; AAA52481.1; JOINED.
DR EMBL; M80810; AAA52481.1; JOINED.
DR EMBL; M80811; AAA52481.1; JOINED.
DR EMBL; M80812; AAA52481.1; JOINED.
DR EMBL; M80813; AAA52481.1; JOINED.
DR EMBL; M80814; AAA52481.1; JOINED.
DR PIR; A33427; HHUHPA.
DR MIM; 230000; .
DR InterPro; IPR000933; Alpha_L_fucos.
DR Pfam; PF01120; Alpha_L_fucos. 1.
DR PRINTS; PR00741; GHYHDLRASE29.
DR PROSITE; PS00385; ALPHA_L_FUCOSIDASE; 1.
KW Hydrolase; Glycosidase; Lysosome; Glycoprotein; Signal; Polymorphism;
KM Disease mutation.
FT SIGNAL 1 22
FT CHAIN 23 461 TISSUE ALPHA-L-FUCOSIDASE.
FT ACT_SITE 291 291 MAY BE IMPORTANT FOR CATALYSIS.
FT CARBOHD 236 236 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHD 263 263 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHD 377 377 N-LINKED (GLCNAC...) (POTENTIAL).
FT CAROHD 60 60 G -> D (IN FUCOSIDOSIS; LOSS OF
ACTIVITY).
FT FTID-VAR_002442.
FT S -> L (IN FUCOSIDOSIS).
FT FTID-VAR_002443.
FT Q -> R.

FT	CONFLICT	71	72	/Frid=VAR_002444.			
FT	CONFLICT	264	264	WE -> FL (IN REF. 3 AND 4).			
FT	CONFLICT	422	422	C -> S (IN REF. 1 AND 2).			
FT	CONFLICT	461	53176	Q -> P (IN REF. 4).			
SO	SEQUENCE	461	AA;	53176	MM;	EAA4AF7119C113CB	CRC64;
Query Match				51.0%;	Score 50;	DB 1;	Length 461;
Best Local Similarity				47.1%;	Pred. No. 1.3;		
Matches				8;	Conservative	1;	Mismatches 8; Indels 0; Gaps 0;
QY	2	PAMMDAESFGAIAHGL	18				
Db	45	PAMFDEAKFGVEIHGWV	61				
RESULT	2						
FUCO_CAEEL	FUCO_CAEEL	STANDARD;	PRT;	407	AA.		
AC	P49713;						
DT	01-OCT-1996 (Rel. 34, Created)						
DT	01-OCT-1996 (Rel. 34, Last sequence update)						
DT	01-OCT-1996 (Rel. 34, Last annotation update)						
DE	Putative alpha-L-fucosidase precursor (EC 3.2.1.51) (Alpha-L-fucoside						
DE	fucosylhydrolase).						
GN	W03611.3.						
OS	Caenorhabditis elegans.						
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;						
OC	Rhabditidae; Peloderinae; Caenorhabditis.						
OX	NCBI_TaxID=6239;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=BRISTOL NZ;						
RA	McMurray A.;						
RL	Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.						
CC	-I- FUNCTION: ALPHA-L-FUCOSIDASE IS RESPONSIBLE FOR HYDROLYZING THE						
CC	ALPHA-1,6-LINKED FUCOSE JOINED TO THE REDUCING-END						
CC	N-ACETYLGLUCOSAMINE OF THE CARBOHYDRATE MOIETIES OF GLYCOPROTEINS						
CC	(BY SIMILARITY)						
CC	-I- CATALYTIC ACTIVITY: An alpha-L-fucoside + H(2)O = an alcohol + L-						
CC	fucose.						
CC	-I- SIMILARITY: BELONGS TO FAMILY 29 OF GLYCOSYL HYDROLASES.						
CC	-----						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration						
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -						
CC	the European Bioinformatics Institute. There are no restrictions on its use						
CC	by non-profit institutions as long as its content is in no way						
CC	modified and this statement is not removed. Usage by and for commercial						
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/						
CC	or send an email to license@isb-sib.ch).						
CC	-----						
DR	EMBL: 267738; GAA91546.1; -.						
DR	InterPro: W03611.3; CE03783.						
DR	InterPro: IPR000933; Alpha_L_fucos.						
DR	Pfam: PF01120; Alpha_L_fucos. 1.						
DR	PRINTS; PR00741; GLHYDRLASE29.						
DR	PROSITE; PS00385; ALPHA_L_FUCOSIDASE; 1.						
KW	Hypothetical protein; Hydrolase; Glycosidase; signal.						
FT	SIGNAL	1	16				
FT	CHAIN	17	407				
FT	ACT_SITE	291	291				
FT	-----						
FT	CARBOHD	197	197				
FT	CARBOHD	358	358				
FT	CARBOHD	374	374				
SO	SEQUENCE	407	AA;	47999	MM;	F28652D614347E20	CRC64;
Query Match				50.0%;	Score 49;	DB 1;	Length 407;
Best Local Similarity				47.1%;	Pred. No. 1.7;		
Matches				8;	Conservative	1;	Mismatches 8; Indels 0; Gaps 0;
QY	2	PAMMDAESFGAIAHGL	18				

Db 31 PSWYDSKFGIFCHWGL 47

RESULT 3
DSBD_PSESP STANDARD: PRT: 578 AA.
ID DSBD_PSESP
AC Q9KJ23;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thiol:disulfide interchange protein dsb precursor.
GN DSBD.
OS Pseudomonas sp.
OC Bacteria; Proteobacteria.
OX NCBI_Taxid=306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JRI / K1;
RA Johann A., Michel J., Averhoff B., Gottschalk G.;
RT "The two-component signal transduction system Arms is involved in
RT deregulation of 3-isopropylcatechol dioxygenase in the mutant strain
RT Pseudomonas sp. K1."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -i- FUNCTION: REQUIRED TO FACILITATE THE FORMATION OF CORRECT
CC DISULFIDE BONDS IN SOME PERIPLASMIC PROTEINS AND FOR THE ASSEMBLY
CC OF THE PERIPLASMIC C-TYPE CYTOCHROMES. ACTS BY TRANSFERRING
CC ELECTRONS FROM CYTOPLASMIC THIOREDOXIN TO THE PERIPLASM. THIS
CC TRANSFER INVOLVES A CASCADE OF DISULFIDE BOND FORMATION AND
CC REDUCTION STEPS (BY SIMILARITY).
CC -i- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -i- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY. DSBD SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF155506; AAF80267.1; -
CC InterPro: IPR000063; Thioredoxin.
CC PROSITE: PS00194; THIOREDOXIN; 1.
CC Redox-active center; Electron transport; Transmembrane;
CC Inner membrane; Cytochrome c-type biogenesis; Signal.
CC SIGNAL 1
FT CHAIN 19 578
FT DOMAIN 19 158
FT TRANSMEM 159 179
FT DOMAIN 180 207
FT TRANSMEM 208 228
FT DOMAIN 229 237
FT TRANSMEM 238 258
FT DOMAIN 259 281
FT TRANSMEM 282 302
FT DOMAIN 303 321
FT TRANSMEM 322 342
FT DOMAIN 343 351
FT TRANSMEM 352 372
FT DOMAIN 373 394
FT TRANSMEM 394 408
FT DOMAIN 409 429
FT TRANSMEM 430 578
FT DISULFID 118 124
FT DISULFID 179 229
FT DISULFID 494
SO SEQUENCE 578 AA; 62181 MW; 7473FC17A9BF7284 CRC64;

Query Match

46.9%; Score 46; DB 1; Length 578;

Best Local Similarity 53.8%; Pred. No. 7.3;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 3 AAMDASFGLAH 15
Db 396 AMKSEGFRAH 408

RESULT 4
FUCCO_RAT STANDARD: PRT: 462 AA.
ID FUCCO_RAT
AC P17164;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Tissue alpha-L-fucosidase precursor (EC 3.2.1.51) (Alpha-L-fucosidase
DE I) (Alpha-L-fucoside fucohydrolase).
GN FUCCAL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 29-40; 90-124 AND 307-372.
RC TISSUE=Liver;
RX MEDLINE=90147522; Pubmed=2482732;
RA Fisher K.J., Aronson N.N., Jr.;
RT "Isolation and sequence analysis of a cDNA encoding rat liver
RT alpha-L-fucosidase."
RL Biochem. J. 264:695-701(1989).
CC -i- FUNCTION: ALPHA-L-FUCOSIDASE IS RESPONSIBLE FOR HYDROLYZING THE
CC ALPHA-1,6-LINKED FUCOSE JOINED TO THE REDUCING-END
CC N-ACETYLGLUCOSAMINE OF THE CARBOHYDRATE MOIETIES OF GLYCOPROTEINS.
CC -i- CATALYTIC ACTIVITY: An alpha-L-fucoside + H(2)O = an alcohol + L-
CC fucose.
CC -i- SUBUNIT: HOMOTETRAMER.
CC -i- SUBCELLULAR LOCATION: Lysosomal.
CC -i- SIMILARITY: BELONGS TO FAMILY 29 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X16145; CAA34268.1; -
CC PIR: S07074; S07074.
CC PIR: S10235; S10235.
CC InterPro: IPR000933; Alpha L. fucos.
CC DR Pfam: PF01120; Alpha L. fucos. 1.
CC DR PRINTS: PR00741; GLYHDLASE29.
CC DR PROSITE: PS00385; ALPHA-L-FUCOSIDASE; 1.
CC KM Hydrolyase; Glycosidase; Lysosome; Glycoprotein; Signal.
CC FT SIGNAL 1 28
CC FT CHAIN 29 462
CC FT ACT_SITE 292 292
CC FT CARBOHYD 237 237
CC FT CARBOHYD 264 264
CC FT CARBOHYD 378 378
SO SEQUENCE 462 AA; 53486 MW; B9B06B62E6019C15 CRC64;

Query Match 45.9%; Score 45; DB 1; Length 462;
Best Local Similarity 41.2%; Pred. No. 8.4;
Matches 7; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

OY 2 PAMDAESFGAIHGL 18
Db 46 PAMDEAKFGIFVHWGV 62

RESULT 5
ID DJB8_MOUSE STANDARD: PRT: 227 AA.
AC O90Y17
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DnaJ homolog subfamily B member 8 (mbJ6).
GN DnaJB8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1;
RX MEDLINE=21023480; PubMed=1147971;
RA Ohtsuka K., Hata M.;
RT "Mammalian HSP40/DNAJ homologs: cloning of novel cDNAs and a proposal
RT for their classification and nomenclature."
RL Cell Stress Chaperones 5:98-112(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Konno H., Aichi J., Fukuda S.,
RA Aikawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Flischiemann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuohi P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
RA Schirni L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Mashio T.,
RA Sawai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.-J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection."
RT Nature 409:685-690(2001).
CC -i SIMILARITY: CONTAINS 1 J DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>).
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL: AB028856: BAA8304.1; -
CC EMBL: AK006026: BAB24372.1; -
CC HSSP: P08622: IBOZ.
DR MGI:1922801: DnaJB8.
DR InterPro: IPR001623: DnaJ_N.
DR Pfam: PF00226: DnaJ_1.
DR SMART: SM00271: DnaJ_1.
DR PROSITE: PS00636: DnaJ_1; 1.
DR PROSITE: PS0076: DnaJ_2; 1.
KW Chaperone.
FT DOMAIN 3 69 J-DOMAIN.
SQ SEQUENCE 227 AA; 25230 MW; 7A6E4CB7989FC095 CRC64;

Query Match	45.4%	Score 44.5;	DB 1;	Length 227;
Best Local Similarity	52.9%	Pred. No. 5.1;		
Matches	9;	Conservative	3;	Mismatches 4;
				Indels 1;
				Gaps 1

```
QY      2 PAWMDA-ESFGAIAHGC 17
          ||::|  || : |||
Db     140 PAFMEALSSFNTLGHGC 156
```

RESULT	6
AC	FUCO_CANFA
ID	FUCO_CANFA
STANDARD:	PRT: 465 AA.
DT	01-FEB-1996 (Rel. 33, Created)
DT	01-FEB-1996 (Rel. 33, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Tissue alpha-L-fucosidase precursor (EC 3.2.1.51) (Alpha-L-fucosidase I) (Alpha-L-fucoside fucosyltransferase).
CN	FUCAL
OS	Canis familiaris (dog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX	NCBI_Taxid:9615;
NP	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-ENGLISH SPRINGER SPANIEL; TISSUE=Liver, and Blood;
RX	MEDLINE=96282899; PubMed=8730282;
RA	Skelly B.J., Sargan D.R., Heritage M.E., Winchester B.G.;
RT	"Isolation of the canine alpha-L-fucosidase cDNA and definition of the fucosidosis mutation in English Springer Spaniels.";
RL	Mamm. Genome 7:271-274(1996).
CC	-1- FUNCTION: ALPHA-L-FUCOSIDASE IS RESPONSIBLE FOR HYDROLYZING THE ALPHA-1,6-LINKED FUCOSE JOINED TO THE REDUCING-END
CC	N-ACETYLGLUCOSAMINE OF THE CARBOHYDRATE MOTIETIES OF GLYCOPROTEINS.
CC	-1- CATALYTIC ACTIVITY: An alpha-L-fucoside + H(2)O = an alcohol + L-fucose.
CC	-1- SUBUNIT: HOMOTETRAMER.
CC	-1- SUBCELLULAR LOCATION: Lysosomal.
CC	-1- DISEASE: A DEFICIENCY OF THIS ACID HYDROLASE IN HUMANS AND DOGS RESULTS IN THE LYSOSOMAL STORAGE DISEASE FUCOSIDOSIS.
CC	-1- SIMILARITY: BELONGS TO FAMILY 29 OF GLYCOSYL HYDROLASES.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-slb.ch/announce/ or send an email to license@isb-slb.ch).
CC	-----
DR	EMBL; X92671; CAA63362.1; -
DR	EMBL; X92672; CAA63362.1; JOINED.
DR	EMBL; X92673; CAA63362.1; JOINED.
DR	EMBL; X92674; CAA63362.1; JOINED.
DR	EMBL; X92675; CAA63362.1; JOINED.
DR	EMBL; X92676; CAA63362.1; JOINED.
DR	EMBL; X92677; CAA63362.1; JOINED.
DR	EMBL; X92678; CAA63362.1; JOINED.
DR	EMBL; X92448; CAA63197.1; -
DR	EMBL; U29765; AAB17403.1; -
DR	EMBL; U29766; AAB17401.1; -
DR	InterPro; IPR000933; Alpha_L_fucos.
DR	Pfam; PF01120; Alpha_L_fucos; 1.
DR	PRINTS; PR00741; GLHYDRLASE29.
DR	PROSITE; PS00385; ALPHA_L_FUCOSIDASE; 1.
KW	Hydrolase; Glycosidase; Lysosome; Glycoprotein; signal.
FT	SIGNAL 1 26
FT	CHAIN 27 465
FT	ACT_SITE 236 296
FT	CARBOHD 241 241
FT	POTENTIAL 241 241
FT	TISSUE ALPHA-L-FUCOSIDASE.
FT	MAY BE IMPORTANT FOR CATALYTIC.
FT	N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 268 268 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 14 15 AV -> LEL (IN REF. 2).
FT CONFLICT 30 32 AAA -> GPP (IN REF. 2).
FT CONFLICT 66 66 K -> V (IN REF. 2).
FT CONFLICT 234 234 K -> E (IN REF. 2).
FT CONFLICT 257 257 L -> G (IN REF. 1); CAA63197).
FT CONFLICT 291 291 L -> H (IN REF. 2).
SQ SEQUENCE 465 AA; 53757 MW; 832BA3C3510341B7 CRC64;

Query Match 44.9%; Score 44; DB 1; Length 465;
Best Local Similarity 43.8%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Oy 2 PAMDAESFGAIAHG 17
Db 50 PDWFDKAKFGVFWHG 65

RESULT 7
SYN_CAEEL STANDARD; PRT; 545 AA.
ID SYN_CAEEL
AC 019722;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Probable asparaginyl-tRNA synthetase, cytoplasmic (EC 6.1.1.22)
DE (Asparagine--tRNA ligase) (Asnrs).
GN F22D6.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Poloderinae; Caenorhabditis.
RN NCBI_TaxID=6239;
RX SEQUENCE FROM N.A.
RP STRAIN-BRISTOL N2;
RA Wilkinson J.;
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-asparagine + tRNA(Asn) = AMP +
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; 271262; CAA95808.1; -;
CC DR HSSP; 052428; 1B8A.
CC DR WormPep; F22D6.3; CE05684.
CC DR InterPro; IPR002106; AA.TRNA.Ligase_II.
CC DR InterPro; IPR002309; tRNA-synt_2.
CC DR InterPro; IPR002312; tRNA-synt_2.
CC DR Pfam; PF00152; tRNA-synt_2; 1.
CC DR Pfam; PF01336; tRNA-anti_1.
CC DR PRINTS; PRO1042; TRNASYNTHASP.
CC DR PROSITE; PS00179; AA.TRNA.LIGASE_II.1; FALSE NEG.
CC DR PROSITE; PS00339; AA.TRNA.LIGASE_II.2; FALSE NEG.
CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
SQ SEQUENCE 545 AA; 61185 MW; 53FADA68ECB81BAC CRC64;

Query Match 44.9%; Score 44; DB 1; Length 545;
Best Local Similarity 42.9%; Pred. No. 14;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
Oy 4 WMDAESFGAIAHG 17
:|:|:|:|:|

Db 499 YMDRRKGYSPVHG 512
RESULT 8
ID DUB6_HUMAN STANDARD; PRT; 326 AA.
AC 075190; 095806; Q9UIK6;
DT 30-MAY-2000 (Rel. 39, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DnaJ homolog subfamily B member 6 (heat shock protein J2) (HSJ-2)
DE (MSJ-1) (HHDJ1) (MRJ).
GN DNAJB6 OR HSJ2 OR MSJ1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN NCBI_TaxID=9606;
RX SEQUENCE FROM N.A. (ISOFORM B).
RP Tissue-Testis;
RA Saito T., Seki N.;
RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM B).
RA Zhang W., Man T., Yuan Z., Cao X.;
RT "HSJ2, a novel human homologue of the bacterial heat-shock protein
RT DnaJ.";
RL Submitted (Jul-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM B).
RC Tissue-Testis;
RX MEDLINE=99115663; PubMed=9915854;
RA "Pituitary tumor-transforming gene protein associates with ribosomal
RT protein S10 and a novel human homologue of DnaJ in testicular cells.";
RL J. Biol. Chem. 274:3151-3158(1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS AND B).
RA Hanai R.;
RT "DnaJ homologue HSJ2b.";
RL Submitted (Jun-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM B).
RA Zhang J.S., Nelson M., Wang L., Smith D.I.;
RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM B).
RC Tissue-Kidney;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weill B., Wellenreuther R., Gassenhuber J., Gassl S.,
RA Ansgore W., Boecker M., Bloeker H., Bauerachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,
RA Lewes H.-W., Ottenweider B., Obermayer B., Tampe J., Heubner D.,
RA Wamplit R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORMS AND B).
RC Tissue-Skin, and Placenta;
RA Strausberg R.;
RL Submitted (Nov-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBUNIT: INTERACTS WITH PTG.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a (shown here) and b, are
CC produced by alternative splicing.
CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -1- CAUTION: Ref. 3 sequence differs from that shown due to a
CC frameshift in position 197.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

^a A 570-kb DNA sequence of the *Escherichia coli* K-12 genome

RA Mueller-Taubenberger A., Westphal M., Noegel A., Gerisch G.;
RT "A developmentally regulated gene product from Dictyostelium

RT discoidium shows high homology to human alpha-L-fucosidase.";
RL FEBS Lett. 246:185-192(1989).
CC -1- FUNCTION: ALPHA-L-FUCOSIDASE IS RESPONSIBLE FOR HYDROLYZING THE
CC ALPHA-1,6-LINKED FUCOSE JOINED TO THE REDUCING-END
CC N-ACETYLGLUCOSAMINE OF THE CARBOHYDRATE MOIETIES OF GLYCOPROTEINS.
CC -1- CATALYTIC ACTIVITY: An alpha-L-fucoside + H₂O = an alcohol + L-
CC fucose.
CC -1- MISCELLANEOUS: THE MAXIMUM EXPRESSION OF ALPHA-L-FUCOSIDASE OCCURS
CC DURING AGGREGATION STAGE OF DICHYSTELIUM DISCOIDEUM.
CC -1- SIMILARITY: BELONGS TO FAMILY 29 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y07497; CAA68800.1; -
DR PIR: A30364; A30364.
DR DictyDb; DD02008; alfa.
DR InterPro; IPR000933; Alpha_L_fucos.
DR Pfam; PF01120; Alpha_L_fucos; 1.
DR PRINTS; PR00741; GLHYDRLASE29.
DR PROSITE; PS00385; ALPHA_L_FUCOSIDASE; 1.
DR Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 461 ALPHA-L-FUCOSIDASE.
FT AC_SITE 288 288 MAY BE IMPORTANT FOR CATALYSIS.
SQ SEQUENCE 461 AA; 52647 MW; CB3D0771FFAB516 CRC64;

Query Match 42.9%; Score 42; DB 1; Length 461;
Best Local Similarity 41.2%; Pred. No. 25;
Matches 7; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

QY 2 PAMNDSEFGAIAHGGL 18
Db 34 PGWYDVVAFGFIHFGI 50

RESULT 11
SYD_PSEAE
ID SYD_PSEAE STANDARD; PRT; 591 AA.
AC OS1422;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Asparaginyl-tRNA synthetase (EC 6.1.1.22) (Aspartate--tRNA ligase)
DE (AspRS).
GN ASPS OR PA0963.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Gabor R.L., Goltzy L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
RN [2]
RP SEQUENCE OF 231-591 FROM N.A.
RC STRAIN-ATCC 15692 / PA01;
RX MEDLINE=97136691; PubMed=8982068;

RA Hishida T., Iwasaki H., Ishioke K., Shinagawa H.;
RT "Molecular analysis of the Pseudomonas aeruginosa genes, ruvA, ruvB
RT and ruvC, involved in processing of homologous recombination
RT intermediates";
RL Gene 182:63-70(1996).
CC -1- CATALYTIC ACTIVITY: ATP + L-aspartate + tRNA(Asp) = AMP +
CC diphosphate + L-asparaginyl-tRNA(Asp).
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE004530; AAG04352.1; -
DR EMBL: D83138; BA011815.1; -
DR HSSP; P21889; 1EQR.
DR InterPro; IPR002106; AA_tRNA_ligase_II.
DR InterPro; IPR004115; GAD.
DR InterPro; IPR002309; tRNA-synt_2.
DR InterPro; IPR002312; tRNA-synt_2.
DR Pfam; PF02938; GAD; 1.
DR Pfam; PF00152; tRNA-synt_2; 2.
DR Pfam; PF01356; tRNA-anti_1.
DR PRINTS; PR01042; TRNASYNTHASP.
DR PROSITE; PS00179; AA_tRNA_LIGASE_II_1; 1.
DR PROSITE; PS00339; AA_tRNA_LIGASE_II_2; 1.
KM Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
FT CONFLICT 588 P -> A (IN REF. 2).
FT SEQUENCE 591 AA; 66207 MW; 62B278CA0DE70A4 CRC64;

Query Match 42.9%; Score 42; DB 1; Length 591;
Best Local Similarity 57.1%; Pred. No. 33;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 MDASEFGAIAHGGL 18
Db 519 LDALIKYCAPPHGGL 532

RESULT 12
SYN_STRPY
ID SYN_STRPY STANDARD; PRT; 448 AA.
AC G9A0R9;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Asparaginyl-tRNA synthetase (EC 6.1.1.22) (Asparagine--tRNA ligase)
DE (AsnRS).
GN ASNS OR SPY0651.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-asparagine + tRNA(Asn) = AMP +
CC diphosphate + L-asparaginyl-tRNA(Asn).
CC -----

CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE006519; AKK3614.1; -
CC DR HSSP: 052428; 1B8A.
CC DR InterPro: IPR002106; AACRNA_LigaseII.
CC DR InterPro: IPR002312; tRNA-synt-asp.
CC DR InterPro: IPR004365; tRNA-anti.
CC DR Pfam: PF01336; tRNA-anti.1.
CC DR PRINTS: PR01042; TRNASYNTHASP.
CC DR PROSITE: PS00179; AA-TRNA_LIGASE_IL1_1.
CC DR PROSITE: PS00339; AA-TRNA_LIGASE_IL2_1.
CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 448 AA; 51207 MW; E19800F0F4F61E97 CRC64;

Query Match 41.8%; Score 41; DB 1; Length 448;
Best Local Similarity 35.7%; Pred. No. 36;
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 4 WMDAESFGAIAHGG 17
DB 402 YLDLRKYGSPHGG 415

RESULT 13
VGIC_BPPZA STANDARD; PRT; 97 AA.
AC P06949;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-JAN-1988 (Rel. 06, Last annotation update)
DE Early protein GPIC.
GN IC.
OS Bacteriophage PZA.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae.
RN NCBI_TaxID=10757;
RP SEQUENCE FROM N.A.
RX MEDLINE=86056991; PubMed=3934048;
RA Paces V., Vlcek C., Urbanek P., Hostomsky Z.;
RT "Nucleotide sequence of the major early region of Bacillus subtilis
RT phage PZA, a close relative of phi 29.";
RL Gene 38:45-56(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M1813; AAA8479.1; -
CC DR PIR: C24528; ERBP12.
KW Early protein.
SQ SEQUENCE 97 AA; 11159 MW; 40FECB8D399FA25 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 97;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 AWMDAESFGAIAHG 16
DB 82 AMVIVLGEVAHG 95

RESULT 14
BCHF_RHOSH STANDARD; PRT; 160 AA.
ID BCHF_RHOSH
AC Q53222;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 2-vinyl bacteriochlorophyllide hydratase (EC 4.2.1.-).
GN BCHF.
OS Rhodospirillum rubrum (Rhodospirillum rubrum) sphaeroides.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum group;
OC Rhodospirillum.
OX NCBI_TaxID=1063;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 8253;
RA Naylor G.W., Adleese H.A., Gibson L.C.D., Hunter C.N.;
RT "The photosynthesis gene cluster of Rhodospirillum sphaeroides";
RT Photosyn. Res. 62:121-139(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RX MEDLINE=20115911; PubMed=10648776;
RA Choudhary M., Kaplan S.;
RT "DNA sequence analysis of the photosynthesis region of Rhodospirillum
RT sphaeroides 2.4.1.";
RT Nucleic Acids Res. 28:862-867(2000).
RN [3]
RP SEQUENCE OF 1-43 FROM N.A.
RC STRAIN=ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158;
RA Gmelinsky M., Kaplan S.;
RT "Genetic evidence that Ppsr from Rhodospirillum sphaeroides 2.4.1
RT functions as a repressor of puc and bchf expression.";
RT Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -1- PATHWAY: Light-independent bacteriochlorophyll biosynthesis.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ010302; CAB38726.1; -
CC DR EMBL: AF195122; AAF24276.1; -
CC DR EMBL: I37197; AAA51373.1; -
CC KW Photosynthesis; Bacteriochlorophyll biosynthesis; Lyase.
SQ SEQUENCE 160 AA; 17984 MW; B2DB511518BD6E24 CRC64;

Query Match 40.8%; Score 40; DB 1; Length 160;
Best Local Similarity 62.5%; Pred. No. 19;
Matches 10; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

QY 1 APA--WMDAESFGAIA 14
DB 94 APAFMEDEVFSFGVIA 109

RESULT 15
BCHF_RHOC A STANDARD; PRT; 171 AA.
AC P26165;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE 2-vinyl bacteriochlorophyllide hydratase (EC 4.2.1.-).

GN BCHF.
OS Rhodobacter capsulatus (Rhodospirillum rubrum capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=1061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB1003;
RX MEDLINE=93224465; PubMed=8385667;
RA Burke D.H., Alberti M., Hearst J.E.;
RT "bchNBNH bacteriochlorophyll synthesis genes of Rhodobacter
RT capsulatus and identification of the third subunit of
RT light-independent protochlorophyllide reductase in bacteria and
RT plants.";
RU J. Bacteriol. 175:2414-2422(1993)
CC -I- PATHWAY: Light-Independent bacteriochlorophyll biosynthesis.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z11165; CAA77527.1; -
DR PIR: S17811; S17811.
DR PIR: A49851; A49851.
KW Photosynthesis; Bacteriochlorophyll biosynthesis; Lyase.
SQ SEQUENCE 171 AA: 19254 MW: 625CE7AA15F3E1C CRC64;

Query Match 40.8%; Score 40; DB 1; Length 171;
Best Local Similarity 62.5%; Pred. No. 20;
Matches 10; Conservative 0; Mismatches 4; Indels 2; Gaps 1;
OY 1 APA-WMDAESFGAIA 14
||| | | |||
DB 94 APAFWEDVFSFGVIA 109

Search completed: July 8, 2002, 08:21:22
Job time: 182 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:10:40 ; Search time 26.89 Seconds

(without alignments)
64,322 Million cell updates/sec

Title: US-09-582-296-1

Perfect score: 98

Sequence: 1 APAMDAESFGAIAHGSL 18

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : PIR_71:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	51.0	467	1	HMHUFA
2	49	50.0	407	2	T26127
3	47	48.0	583	2	G81412
4	46	46.9	400	2	T46383
5	45.5	46.4	352	2	B85518
6	45.5	46.4	378	2	F90667
7	45	45.9	224	2	AG0406
8	45	45.9	363	2	A83563
9	45	45.9	449	2	H87019
10	45	45.9	444	2	G72393
11	45	45.9	462	2	S10235
12	44	44.9	517	2	D86637
13	44	44.9	545	2	T21253
14	43.5	44.4	258	2	A69830
15	43	43.9	242	2	A11883
16	42.5	43.4	769	2	B87681
17	42.5	43.4	808	2	F90915
18	42.5	43.4	808	2	F64914
19	42.5	43.4	808	2	C85764
20	42	42.9	334	2	E69361
21	42	42.9	361	2	PC4295
22	42	42.9	461	1	A30364
23	42	42.9	483	2	F75360
24	42	42.9	541	2	T48811
25	42	42.9	591	2	E83524
26	42	42.9	700	2	I51335
27	42	42.9	1486	2	F86785
28	41.5	42.3	606	2	T35378
29	41	41.8	76	2	E87562

ALIGNMENTS

30	41	41.8	81	2	A90488	alpha-fucosidase N
31	41	41.8	414	2	A83745	N-carbamyl-L-amino
32	41	41.8	419	2	G83355	hypothetical prote
33	41	41.8	472	2	AG3600	cellulose synthase
34	41	41.8	498	2	F82651	hypothetical prote
35	41	41.8	589	2	H82627	aspartyl-tRNA synt
36	41	41.8	644	2	C75457	alpha-amylase - De
37	41	41.8	877	2	A83437	probable ClpA/B-ty
38	41	41.8	1527	2	H70655	glutamate synthase
39	41	41.8	1527	2	E86916	hypothetical prote
40	41	41.8	1530	2	E82085	glutamate synthase
41	40.5	41.3	242	2	F87640	exsb protein (limp
42	40.5	41.3	474	2	S65763	chitinase (EC 3.2.
43	40	40.8	97	1	ERBP12	gene 1C protein -
44	40	40.8	109	2	C75409	conserved hypothet
45	40	40.8	148	2	H71007	hypothetical prote

RESULT 1

HMHUFA
alpha-L-fucosidase (EC 3.2.1.51) 1 precursor, tissue - human

N:Alternate names: alpha-L-fucosidase I; alpha-L-fucoside fucosylolase

C:Species: Homo sapiens (man)

C>Date: 30-Jun-1987 #sequence.revision 30-Jun-1991 #text.change 21-Jul-2000

C:Accession: A33427; A00901; A61016; I54348; I37394; I84457

R:Ochiodoro, T.; Beckmann, K.R.; Morris, C.P.; Hopwood, J.J.

R:Biochem. Biophys. Res. Commun. 164, 439-445, 1989

A:Title: Human alpha-L-fucosidase: complete coding sequence from cDNA clones.

A:Reference number: A33427; MUID:90026416

A:Accession: A33427

A:Molecule type: mRNA

A:Residues: 1-461 <OCC>

A:Cross-references: GB:M29877; NID:g178408; PIDN:AAA3519.1; PID:g178409

R:Fukushima, H.; de Wet, J.R.; O'Brien, J.S.

Proc. Natl. Acad. Sci. U.S.A. 82, 1262-1265, 1985

A:Title: Molecular cloning of a cDNA for human alpha-L-fucosidase.

A:Reference number: A00901; MUID:85140268

A:Accession: A00901

A:Molecule type: mRNA

A:Residues: 73-263, 'C', 265-419 <FSI>

A:Cross-references: GB:M10355

R:Fukushima, H.; Nishimoto, J.; Okada, S.

J. Inher. Metab. Dis. 13, 761-765, 1990

A:Title: Sequencing and expression of a full-length cDNA for human alpha-L-fucosidase

A:Reference number: A61016; MUID:91061435

A:Accession: A61016

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-68, 'T', 70-87 <FS2>

R:Seo, H.C.; Willems, P.J.; Kretz, K.A.; Martin, B.M.; O'Brien, J.S.

Hum. Mol. Genet. 2, 423-429, 1993

A:Title: Fucosidosis: four new mutations and a new polymorphism.

A:Reference number: I54348; MUID:93278392

A:Accession: I54348

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-263, 'C', 265-461 <RES>

A:Cross-references: GB:M60815; NID:g182786; PIDN:AAA52481.1; PID:g182788

R:de Wet, J.R.; Fukushima, H.; Dewji, N.N.; Wilcox, E.; O'Brien, J.S.; Hellinski, D.R.

DNA 3, 437-447, 1984

A:Title: Chromogenic immunodetection of human serum albumin and alpha-L-fucosidase cl

A:Reference number: I37394; MUID:85076174

A:Accession: I37394

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA

A:Residues: 357-393 <RE2>

A:Cross-references: EMBL:X01390; NID:g31475; PIDN:CAA25646.1; PID:g1335066

A:Accession: I84457

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: mRNA
A:Residues: 70,'FL','73-263,'C','265-421,'P' <RE3>
A:Cross-references: GB:M10157; NID:q182778; PID:q182779
C:Comment: The fucosidases hydrolyze alpha-L-fucose from glycolipids and oligosaccharide
C:Genetics:
A:Gene: GDB:FUCAL1
A:Cross-references: GDB:119237; OMIM:230000
A:Map position: 1p35-1p34
A:Insertions: 125/2; 170/2; 216/2; 251/3; 318/3; 382/2; 415/3
C:Superfamily: alpha-L-fucosidase
C:Keywords: fucosidosis; glycoprotein; glycosidase; homotetramer; hydrolase; lysosome
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-461/Product: alpha-L-fucosidase #status predicted <MAT>
F:236,246,263,377/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 51.0%; Score 50; DB 1; Length 461;
Best Local Similarity 47.1%; Pred. No. 2.3;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 2 PAMDAESFGAIAHGCL 18
||| ||| ||| |||
Db 45 PAMFDEAKFGVFIHMGV 61

RESULT 2
T26127
hypothetical protein W03G11.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C:Accession: T26127
R:McMurray, A.
Submitted to the EMBL Data Library, November 1995
A:Reference number: Z20156
A:Accession: T26127
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-407 <MIL>
A:Cross-references: EMBL:Z67738; PIDN:CAA91546.1; GSPDB:GN00028; CESP:W03G11.3
A:Experimental source: clone W03G11
C:Genetics:
A:Gene: CESP:W03G11.3
A:Map position: X
A:Insertions: 19/3; 49/3; 127/2; 203/3; 248/2; 355/2
C:Superfamily: alpha-L-fucosidase

Query Match 50.0%; Score 49; DB 2; Length 407;
Best Local Similarity 47.1%; Pred. No. 2.9;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 2 PAMWDAESFGAIAHGCL 18
||| ||| ||| |||
Db 31 PSWYDSDKFGIFCHMGL 47

RESULT 3
G81412
aspartate--tRNA ligase (EC 6.1.1.12) Cj0640c [imported] - Campylobacter jejuni (strain N
C:Species: Campylobacter jejuni
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000
C:Accession: G81412
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barr
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: A81250; MUID:20150912
A:Accession: G81412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-583 <PAR>
A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB75276.1; PID:g696810
A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics:
A:Gene: asps; Cj0640c
A:Superfamily: lysine--tRNA ligase
C:Keywords: ligase

Query Match 48.0%; Score 47; DB 2; Length 583;
Best Local Similarity 64.3%; Pred. No. 9.5;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 MDAESFGAIAHGCL 18
||| ||| ||| |||
Db 513 LDALSFQAPPFHGCI 526

RESULT 4
T46383
hypothetical protein DKFZp43401519.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46383
R:Ottenwelder, B.; Obermer, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
Submitted to the Protein Sequence Database, January 2000
A:Reference number: Z23031
A:Accession: T46383
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-400 <AA>
A:Cross-references: EMBL:AL137540
A:Experimental source: adult testis; clone DKFZp43401519
C:Genetics:
A:Note: DKFZp43401519.1

Query Match 46.9%; Score 46; DB 2; Length 400;
Best Local Similarity 50.0%; Pred. No. 9;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 4 WMDAESFGAIAHGCL 17
||| ||| ||| |||
Db 263 WEDAQCFSLHLSG 276

RESULT 5
B85518
hypothetical protein Z0347 [imported] - Escherichia coli (strain O157:H7, substrain E
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85518
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Cross-references: GB:AF005174; NID:q12513064; PIDN:AAG54606.1; GSPDB:GN00145; UWCP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0347

Query Match 46.4%; Score 45.5; DB 2; Length 352;
Best Local Similarity 64.3%; Pred. No. 9.5;
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

OY 4 WMDAESFGAIAHGCL 17
||| ||| ||| |||
Db 201 WKDAQS-GTFAHGC 213

RESULT 6
F90667
hypothetical protein ECs0310 [imported] - Escherichia coli (strain O157:H7, substrain R1)
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F90667
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kuwakawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90667
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-378 <NAV>
A:Cross-references: GB:BA000007; PIDN:BA033733.1; PID:q13359767; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs0310

Query Match 46.4%; Score 45.5; DB 2; Length 378;
Best Local Similarity 64.3%; Pred. No. 10;
Matches 9; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

OY 4 WMADESFCAIAGG 17
| | | | |
Db 227 WKDAQS-GTFAHGG 239

RESULT 7
AG0406
conserved hypothetical protein YPO3349 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AG0406
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0406
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-224 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92579.1; PID:q15981276; GSPDB:GN00175
C:Genetics:
A:Gene: YPO3349

Query Match 45.9%; Score 45; DB 2; Length 224;
Best Local Similarity 64.3%; Pred. No. 6.8;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 MDAESFGAIAHGL 18
| | | | |
Db 111 MDAVAFGAMKGL 124

RESULT 8
AB3563
conserved hypothetical protein PA0666 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: AB3563
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: AB3563

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-363 <STO>
A:Cross-references: GB:AE004501; GB:AE004091; NID:g9946537; PIDN:AA04055.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0666
C:Superfamily: hypothetical protein H10753

Query Match 45.9%; Score 45; DB 2; Length 363;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 PAMMAESFGAIAH 15
| | | | |
Db 322 PAMMEGMAFAMLAH 335

RESULT 9
H87019
probable glycosyl transferase, ML0886 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87019
R:Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;
R.; Davies, R.M.; Devlin, K.; Dutfoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holro
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.;
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H87019
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-384 <STO>
A:Cross-references: GB:AL450380; NID:q13092958; PIDN:CAC31267.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML0886

Query Match 45.9%; Score 45; DB 2; Length 384;
Best Local Similarity 47.1%; Pred. No. 13;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 1 APAMMAESFGAIAHGG 17
| | | | |
Db 41 APOMTGADTFDMAHGG 57

RESULT 10
G72393
hypothetical protein TM0306 - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: G72393
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garret, M.M.; Stewart, A.M.; Cotton, M.D.; Pratl, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: G72393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-449 <ARN>
A:Cross-references: GB:AE001712; GB:AE000512; NID:g4980799; PIDN:AAD35394.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM0306
C:Superfamily: alpha-L-fucosidase

Query Match 45.9%; Score 45; DB 2; Length 449;
Best Local Similarity 41.2%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
QY 2 PAMMDAESFGAIAHGCL 18
DB 21 PKMFDAKFGIRIHWGI 37

RESULT 11
alpha-L-fucosidase (EC 3.2.1.51) precursor - rat
S10235
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C:Accession: S07074; S10235
R:Fisher, K.J.; Aronson Jr., N.N.
Biochem. J. 264, 695-701, 1989
A:Title: Isolation and sequence analysis of a cDNA encoding rat liver alpha-L-fucosidase
A:Reference number: S07074; MUID:90147522
A:Accession: S07074
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-376; 'R', 378-462 <FIS>
A:Cross-references: EMBL:X16145; NID:955650; PIDN:CAA34268.1; PID:955651
R:Aronson Jr., N.N.
Submitted to the EMBL Data Library, August 1989
A:Reference number: S10235
A:Accession: S10235
A:Molecule type: mRNA
A:Residues: 1-462 <ARO>
A:Cross-references: EMBL:X16145
C:Superfamily: alpha-L-fucosidase
C:Keywords: glycosidase; hydrolase

Query Match 45.9%; Score 45; DB 2; Length 462;
Best Local Similarity 41.2%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
QY 2 PAMMDAESFGAIAHGCL 18
DB 46 PRWFDEAKFGLEFHWGV 62

RESULT 12
ABC transporter ATP binding protein ybab [imported] - Lactococcus lactis subsp. lactis
D86637
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: D86637
R:Boletín, A.; Winkler, P.; Mauger, S.; Jallón, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s.s.
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: D86637
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-517 <STO>
A:Cross-references: GB:AEO05176; PID:912722945; PIDN:AAK04198.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: ybab
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 44.9%; Score 44; DB 2; Length 517;
Best Local Similarity 52.9%; Pred. No. 26;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 2 PAMMDAESFGAIAHGCL 18
DB 412 PSMWQALRGALARVGL 428

RESULT 13
hypothetical protein F22D6.3 - Caenorhabditis elegans
T21253
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T21253
R:Wilkinson, J.
Submitted to the EMBL Data Library, April 1996
A:Reference number: Z19397
A:Accession: T21253
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-545 <NTL>
A:Cross-references: EMBL:T21262; PIDN:CAA95808.1; GSPDB:GN00019; CESP:F22D6.3
A:Experimental source: clone F22D6
C:Genetics:
A:Gene: CESP:F22D6.3
A:Map position: 1
A:Introns: 28/3; 160/3; 318/2; 376/3; 430/3
C:Superfamily: lysine--tRNA ligase

Query Match 44.9%; Score 44; DB 2; Length 545;
Best Local Similarity 42.9%; Pred. No. 28;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 4 WMDAESFGAIAHGCL 17
DB 499 YMDORKYGVPHGG 512

RESULT 14
hypothetical protein yhfC - Bacillus subtilis
A69830
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C:Accession: A69830
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
A.; Ehrlich, S.D.; Emerson, P.T.; Eutlian, K.D.; Eittington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funo, S.; Gallazzi, A.; Gal
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogg, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivoita, C.; Rocha, E.; Koche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchida
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
A:Reference number: A69830; MUID:98044033
A:Accession: A69830
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-258 <KUN>
A:Cross-references: GB:Z99109; GB:AL009126; NID:92633260; PIDN:CAB12658.1; PID:el1830
A:Experimental source: strain 168
C:Genetics:
A:Gene: yhfC

Query Match 44.4%; Score 43.5; DB 2; Length 258;
Best Local Similarity 60.0%; Pred. No. 14;
Matches 9; Conservative 1; Mismatches 4; Indels 1; Gaps 1;
QY 4 WMDAESFGAIAHGCL 18
DB 108 WADGLAFGA-GHGCL 121

RESULT 15

AI1883

hypothetical protein alr0618 [imported] - *Anabaena* sp. (strain PCC 7120)

C; Species: *Anabaena* sp.

A/Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence revision 14-Dec-2001 #text change 11-Jan-2002

C:\Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:\Accession: A11893

C;Accession: A11883

R; Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuriltz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 9: 205-213, 2001

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*

A; Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: A11883

A;Accession: A11883
A;Status: preliminary

A:status: preliminary
A:molecule type: DNA

A; molecule type: DNA
A; Residues: 1-242 <KIR>

A;Residues: 1-242 <KUR>

A; Cross-references: GB:BA000019; PIDN:

A; Experiment

A;Gene: a1r0618

Query Match

43.98; Score 43; DB 2; Length 242;

Best Local Similarity 61.5%; Pred. No. 16;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 MDAESFGAIAHGG 17

MIKE: I'm not a doctor, but I think you should see a doctor.

Db 218 IDAEKI,NAIAYGG 230

Search completed: July 8, 2002, 08:20:15
Job time: 575 sec

Job time: 575 sec

US-08-932-376A-4
; Sequence 4, Application US/08932376A
; Patent No. 5869309
; GENERAL INFORMATION:
; APPLICANT: Politino, Michael
; APPLICANT: Tonzi, Sean M.
; APPLICANT: Usher, John J.
; APPLICANT: Burnett K, William V.
; APPLICANT: Romancik, Guna
; TITLE OF INVENTION: CEPHALOSPORIN ESTERASE GENE FROM
; TITLE OF INVENTION: RHODOSPORIDIUM TORULOIDES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bristol-Myers Squibb Company
; STREET: Rt. 206 & Provinceline Road
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,376A
; FILING DATE: 17-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Savitsky, Thomas R.
; REGISTRATION NUMBER: 31,661
; REFERENCE/DOCKET NUMBER: ON0144a
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-4956
; TELEFAX: (609) 252-4526
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 544 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-932-376A-4

Query Match 49.0%; Score 48; DB 2; Length 544;
Best Local Similarity 52.9%; Pred. No. 4;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAMMDAESFGAIAHGCL 18
| : ||| : |||
DB 414 PTYWTAEAFGSSAHKGL 430

RESULT 3
US-08-932-376A-2
; Sequence 2, Application US/08932376A
; Patent No. 5869309
; GENERAL INFORMATION:
; APPLICANT: Politino, Michael
; APPLICANT: Tonzi, Sean M.
; APPLICANT: Usher, John J.
; APPLICANT: Burnett K, William V.
; APPLICANT: Romancik, Guna
; TITLE OF INVENTION: CEPHALOSPORIN ESTERASE GENE FROM
; TITLE OF INVENTION: RHODOSPORIDIUM TORULOIDES
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bristol-Myers Squibb Company
; STREET: Rt. 206 & Provinceline Road
; CITY: Princeton
; STATE: New Jersey
; COUNTRY: USA

ZIP: 08543-4000
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,376A
; FILING DATE: 17-SEP-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Savitsky, Thomas R.
; REGISTRATION NUMBER: 31,661
; REFERENCE/DOCKET NUMBER: ON0144a
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (609) 252-4956
; TELEFAX: (609) 252-4526
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 572 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-932-376A-2

Query Match 49.0%; Score 48; DB 2; Length 572;
Best Local Similarity 52.9%; Pred. No. 4.2;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 PAMMDAESFGAIAHGCL 18
| : ||| : |||
DB 442 PTYWTAEAFGSSAHKGL 458

US-09-199-637A-202
; Sequence 4
; Sequence 202, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 202
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-199-637A-202

Query Match 46.9%; Score 46; DB 4; Length 550;
Best Local Similarity 53.3%; Pred. No. 8.6;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAMMDAESFGAIAH 15
| | | | : ||| : |||
DB 34 AEAWGAAAYGALAH 48

RESULT 5
US-09-199-637A-201
; Sequence 201, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 201
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-201

Query Match 46.9%; Score 46; DB 4; Length 608;
Best Local Similarity 53.3%; Pred. No. 9,6;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
OY 1 APAMDAESFGAIAH 15
| | | | : | : | : | : |
Db 92 AEAMSAAGALAH 106
RESULT 6
US-09-199-637A-200
; Sequence 200, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 200
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-200

Query Match 46.9%; Score 46; DB 4; Length 639;
Best Local Similarity 53.3%; Pred. No. 10;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
OY 1 APAMDAESFGAIAH 15
| | | | : | : | : | : |

Db 123 AEAMSAAGALAH 137
RESULT 7
US-09-199-637A-103
; Sequence 103, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 103
; LENGTH: 641
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-103

Query Match 46.9%; Score 46; DB 4; Length 641;
Best Local Similarity 53.3%; Pred. No. 10;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
OY 1 APAMDAESFGAIAH 15
| | | | : | : | : | : |
Db 129 AEAMSAAGALAH 143

RESULT 8
US-09-199-637A-199
; Sequence 199, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; CURRENT FILING DATE: 1998-11-25
; PRIOR APPLICATION NUMBER: 60/066,517
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 199
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-199

Query Match 46.9%; Score 46; DB 4; Length 643;
Best Local Similarity 53.3%; Pred. No. 10;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 APAMMDAESFGAIAH 15
||| | :||:|
Db 127 AEWSAAMAYGALAH 141

RESULT 9

US-09-199-637A-253
; Sequence 253, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Raffae, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; FILE REFERENCE: 00786/361002
; CURRENT APPLICATION NUMBER: US/09/199,637A
; PRIOR FILING DATE: 1998-11-23
; PRIOR FILING DATE: 1997-11-25
; NUMBER OF SEQ ID NOS: 437
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 253
; LENGTH: 645
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-253

Query Match 46.9%; Score 46; DB 4; Length 645;
Best Local Similarity 53.3%; Pred. No. 10;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
OY 1 APAMMDAESFGAIAH 15
||| | :||:|
Db 129 AEWSAAMAYGALAH 143

RESULT 10

US-08-868-288A-3
; Sequence 3, Application US/08868288A
; Patent No. 5922567
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/868,288A
; FILING DATE: June 3, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2RAT01
; CLONE: 260873
US-08-868-288A-3

Query Match 43.9%; Score 43; DB 2; Length 330;
Best Local Similarity 70.0%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 9 SEGAIAHGCL 18
|||:|
Db 163 SFGSLGHGCL 172

RESULT 11

US-09-235-373-3
; Sequence 3, Application US/09235373
; Patent No. 6001598
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,373
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE: June 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2RAT01
; CLONE: 260873
US-09-235-373-3

Query Match 43.9%; Score 43; DB 3; Length 330;
Best Local Similarity 70.0%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 9 SFGAIAHGSL 18
|||||
Db 163 SFGSLGHGSL 172

RESULT 12
US-09-388-993-3
Sequence 3, Application US/09388993
Patent No. 6043222

GENERAL INFORMATION:

APPLICANT: Au-Young, Janice

APPLICANT: Lal, Preeti

APPLICANT: Bandman, Olga

TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/388,993

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/868,288

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0309 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 330 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: HMT2RAT01

CLONE: 260873

US-09-388-993-3

Query Match 43.9%; Score 43; DB 3; Length 330;
Best Local Similarity 70.0%; Pred. No. 15;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 9 SFGAIAHGSL 18
|||||
Db 163 SFGSLGHGSL 172

RESULT 13
US-08-598-873-58

Sequence 58, Application US/08598873

Patent No. 5928884

GENERAL INFORMATION:

APPLICANT: Croce, Carlo M.

APPLICANT: Huebner, Kay

TITLE OF INVENTION: FHIT PROTEINS AND NUCLEIC ACIDS AND
METHODS BASED THEREON
NUMBER OF SEQUENCES: 77
CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/598,873

FILING DATE: 09-FEB-1996

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Friedel, Thomas E.

REGISTRATION NUMBER: 29,258

REFERENCE/DOCKET NUMBER: 8666-004

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:

LENGTH: 122 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-598-873-58

Query Match 40.8%; Score 40; DB 2; Length 122;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 APAMDAESF 10
|||||
Db 75 SPWQDAQSF 84

RESULT 14
US-08-605-430-58
Sequence 58, Application US/08605430
Patent No. 6242212

GENERAL INFORMATION:

APPLICANT: Croce, Carlo M.

APPLICANT: Huebner, Kay

TITLE OF INVENTION: FHIT PROTEINS AND NUCLEIC ACIDS AND

METHODS BASED THEREON

NUMBER OF SEQUENCES: 86

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/605,430

FILING DATE: 22-FEB-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Friedel, Thomas E.

REGISTRATION NUMBER: 29,258
REFERENCE/DOCKET NUMBER: 8666-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 780-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 122 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-603-430-58

Query Match 40.8%; Score 40; DB 4; Length 122;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 APAMDAESF 10
: | | | | |
Db 75 SPQWQDAQSF 84

RESULT 15
US-08-844-154-2
Sequence 2, Application US/08844154
Patent No. 5827708
GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 5827708e1 Compounds
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Smithkline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,154
FILING DATE: 18-APR-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9607993.4
FILING DATE: 18-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31457-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 434 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-844-154-2

Query Match 40.8%; Score 40; DB 2; Length 434;
Best Local Similarity 35.7%; Pred. No. 64;
Matches 5; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 4 WMDAESGAIHGC 17
: | | | | |
Db 401 YLDLRKRYGTVPHEG 414

Search completed: July 8, 2002, 08:19:40
Job time: 585 sec

Mon Jul 8 09:33:13 2002

us-09-582-296-1.rai

1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:09:35 ; Search time 51.87 Seconds
(without alignments)
38.545 Million cell updates/sec

Title: us-09-582-296-1

Perfect score: 98
Sequence: 1 APAMMDSEFGAIAHGCL 18

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq.032802:*

- 1: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2000.DAT:*
- 21: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT:*
- 22: /SIDSI/gcgdata/hold-geneseq/geneeqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	98	100.0	18 20 AAY27010	N-terminal fragmen
2	69	70.4	20 20 AAY27011	N-terminal fragmen
3	50	51.0	461 15 AAR56285	Sequence of human
4	49	50.0	129 22 AAB85247	Human secreted pro
5	49	50.0	184 21 AAB14315	Human secreted pro
6	49	50.0	22 AAB85237	Human secreted pro
7	49	50.0	167 19 AAW75057	Human secreted pro
8	49	50.0	20 AAY13578	Amino acid sequenc
9	49	50.0	467 21 AAY94860	Human protein cion
10	49	50.0	467 22 AAU29031	Human PRO polypept
11	49	50.0	467 22 AAW78342	Human protein SEQ

12	49	50.0	467 22 AAB8401	Human membrane or
13	49	50.0	467 22 AAB80246	Human PRO260 prote
14	49	50.0	496 22 AAM79326	Human protein SEQ
15	48	49.0	572 19 AAM54139	Rhodospiridium tor
16	48	49.0	572 22 AAB82871	Rhodospiridium tor
17	46	46.9	296 21 AAY91613	Human secreted pro
18	46	46.9	334 22 AAU61259	Propionibacterium
19	46	46.9	383 22 ABB11745	Human netrin 4 hom
20	46	46.9	549 20 AAY29270	Protein encoded by
21	46	46.9	607 20 AAY29269	Protein encoded by
22	46	46.9	628 20 AAG66449	Human beta-netrin.
23	46	46.9	639 20 AAY29268	Protein encoded by
24	46	46.9	642 20 AAY29266	Virulence factor p
25	46	46.9	643 20 AAY29267	Protein encoded by
26	46	46.9	645 20 AAY29172	Amino acid sequenc
27	45	45.9	122 22 AAU29615	Novel human secret
28	44	44.9	608 22 AAG91551	C glutamicum prote
29	43	43.9	317 20 AAY74126	Human prostate tum
30	43	43.9	330 20 AAM94066	Human DnaJ-like pr
31	43	43.9	468 22 AAM8945	Human novel polypeptide
32	43	43.9	468 22 AAU14130	Human novel protei
33	43	43.9	607 22 AAU63535	Propionibacterium
34	42.5	43.4	280 22 AAG90751	C glutamicum prote
35	42.5	43.4	1222 22 ABB67050	Drosophila melanog
36	42.5	43.4	2506 22 ABB67191	Novel human diagno
37	42.5	43.4	2506 22 ABB67191	Novel human diagno
38	42	42.9	74 22 AAU42894	Novel human diagno
39	42	42.9	591 22 AAU76222	Pseudomonas aerugi
40	42	42.9	700 22 AAE02420	Frog vasa protein.
41	42	42.9	899 22 ABB65488	Drosophila melanog
42	42	42.9	899 22 ABB65489	Drosophila melanog
43	41	41.8	93 21 AAB40640	Human ORFX ORF404
44	41	41.8	111 22 ABB23358	Novel human diagno
45	41	41.8	121 22 AA003564	Human polypeptide

ALIGNMENTS

RESULT 1	
AAV27010	standard; peptide: 18 AA.
ID	AAV27010;
AC	AAV27010;
XX	
DT	24-SEP-1999 (first entry)
XX	
DE	N-terminal fragment of anti-freeze protein from Lichen.
XX	
KW	Antifreeze protein; AFP; Lichen; frozen food product; pasteurisation;
KW	sterilisation; freezing; frozen confectionery product; N-terminal.
OS	Umbilicaria antarctica.
XX	
PN	WO9937673-A2.
XX	
PD	29-JUL-1999.
XX	
PF	23-DEC-1998; 98WO-EP08554.
XX	
PR	22-JAN-1998; 98GB-0001420.
XX	
PA	(UNTL) UNILEVER NV.
PA	(UNTL) UNILEVER PLC.
PI	Byass LJ, Sidebottom CM, Smallwood MF;
DR	WPI, 1999-444595/37.
XX	
PT	New isolated antifreeze protein obtained from Lichen, used for the
PT	preparation of food products, particularly frozen confectionery
PT	products
XX	

PS Claim 1; Page 17; 20pp; English.

CC The invention describes a novel antifreeze protein (AFP) obtained from
CC Lichen that comprises an apparent mol. wt. of 20-28kD and has an
CC N-terminal amino acid sequence which shows at least 80% overlap with
CC the present sequence; and modified versions and isoforms of this
CC protein. The AFP can conveniently be used in food products, preferably
CC in food products which are frozen or intended to be frozen. Especially
CC preferred is the use of AFPs in products which are heated e.g. by
CC pasteurisation or sterilisation prior to freezing and in frozen
CC confectionery products. Using the AFP ingredient, mixes can be frozen
CC under quiescent conditions, e.g. in a shop or home freezer without the
CC formation of unacceptable ice crystal shapes and hence with a texture
CC different to products normally obtained via quiescent freezing. The
CC present sequence represents a N-terminal fragment of an AFP.

CC Sequence 18 AA;

Query Match 100.0%; Score 98; DB 20; Length 18;

Best Local Similarity 100.0%; Pred. No. 1.4e-09; Mismatches 0; Indels 0; Gaps 0;

Db 1 APAMDAESFGATAHGGL 18
1 apavmdaesfgalahggl 18

RESULT 2

AAV27011
ID AAV27011 standard; peptide; 20 AA.

AC AAV27011;

DT 24-SEP-1999 (first entry)

DE N-terminal fragment of anti-freeze protein from Lichen.

KW Antifreeze protein; AFP; Lichen; frozen food product; pasteurisation;
KW sterilisation; freezing; frozen confectionery product; N-terminal.

OS Umbilicaria antarctica.

PN W09937673-A2.

PD 29-JUL-1999.

PE 23-DEC-1998; 98MO-EP08554.

PR 22-JAN-1998; 98GB-0001420.

PA (UNIL) UNILEVER NV.
PA (UNIL) UNILEVER PLC.

PI Byass LJ, Sidebottom CM, Smallwood MF;

DR WPI; 1999-444595/37.

PT New isolated antifreeze protein obtained from Lichen, used for the
PT preparation of food products, particularly frozen confectionery
PT products

PS Claim 2; Page 17; 20pp; English.

CC The invention describes a novel antifreeze protein (AFP) obtained from
CC Lichen that comprises an apparent mol. wt. of 20-28kD and has an
CC N-terminal amino acid sequence which shows at least 80% overlap with
CC sequence shown in AAV27010; and modified versions and isoforms of
CC this protein. The AFP can conveniently be used in food products,
CC preferably in food products which are frozen or intended to be frozen.
CC Especially preferred is the use of AFPs in products which are heated e.g.
CC by pasteurisation or sterilisation prior to freezing and in frozen
CC confectionery products. Using the AFP ingredient, mixes can be frozen

CC under quiescent conditions, e.g. in a shop or home freezer without the
CC formation of unacceptable ice crystal shapes and hence with a texture
CC different to products normally obtained via quiescent freezing. The
CC present sequence represents a N-terminal fragment of an 24 kDa AFP.

CC Sequence 20 AA;

Query Match 70.4%; Score 69; DB 20; Length 20;

Best Local Similarity 80.0%; Pred. No. 0.0001; Mismatches 1; Indels 2; Gaps 1;

Db 1 APAM--DAESFGATAHGGL 18
1 apavmmdaesfgalahggl 20

RESULT 3

AAR56285
ID AAR56285 standard; Protein; 461 AA.

AC AAR56285;

DT 03-MAR-1995 (first entry)

DE Sequence of human alpha-L-fucosidase including signal peptide.

KW Anti-adhesion enzyme; alpha-L-fucosidase.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..22 /label= signal

FT Protein 23..461 /label= mature

FT WO9414472-A.

PD 07-JUL-1994.

PF 21-DEC-1993; 93WO-US12464.

PR 22-DEC-1992; 92US-0994650.

PA (REGC) UNIV CALIFORNIA.

PI Carson DA, Wicks I;

DR WPI; 1994-234351/28.

DR N-PSDB; AA06537.

PT Inhibiting adhesion of cells to selectin-contg. receptors - by
PT treatment with enzyme degrading carbohydrate residues in selectin
PT adhesion ligands, partic. for treating inflammation

PS Disclosure; Page 27-30; 42pp; English.

CC The AA sequence in R56285 is derived from sequence information obtd.
CC through isolation of several human alpha-L-fucosidase cDNA clones
CC reported by Occhiodoro et al. at Biochem. Biophys. Res. Comm.,
CC 164:439-445 (1989). The full-length nucleotide sequence for alpha-L-
CC fucosidase is available from the GENBANK molecular sequence database
CC under Accession No. M28099. The DNA sequence in Q6537 is described
CC as genomic.

CC Sequence 461 AA;

Query Match 51.0%; Score 50; DB 15; Length 461;

Best Local Similarity 47.1%; Pred. No. 5.3; Mismatches 8; Indels 0; Gaps 0;

OS	Homo sapiens.
XX	
PN	WO200144267-A2.
XX	
PD	21-JUN-2001.
XX	
PF	18-DEC-2000; 2000WO-US34433.
XX	
PR	16-DEC-1999; 99US-0465512. 17-DEC-1999; 99WO-US29854.
PA	(CURA-) CURAGEN CORP.
XX	
PI	Shinkets RA;
XX	
DR	WPI: 2001-408469/43.
N-PSDB:	AAM22700.
XX	
PT	New secreted human protein, SECX is useful for the treatment of pathological disorders i.e. cancer -
XX	
PS	Claim 17; Fig 16A; 162pp; English.
XX	
CC	The invention provides novel secreted human proteins (SECX) and nucleotide sequences encoding the SECX polypeptides. The SECX polypeptides can be expressed by standard recombinant methodology. SECX polypeptides are useful for the treatment, prevention or delay of onset of a pathological condition associated with aberrant SECX expression or activity, i.e. cancer. The present sequence represents the amino acid sequence of a human SECX polypeptide (see DE file for clone identification).
XX	
SQ	Sequence 184 AA;
XX	
Query Match	50.0%; Score 49; DB 22; Length 184;
Best Local Similarity	47.1%; Pred No. 2.7;
Matches 8; Conservative 1; Mismatches 8; Indels 0; Gaps 0	
OY	2 PAMWAESFGAIHAGGL 18 Db 47 pawfdqakfglfihgvy 63
RESULT 7	
AAW75057	
ID AAW75057	standard; Protein; 467 AA.
XX	
AC AAW75057;	
XX	
DT 28-JAN-1999	(first entry)
XX	
DE Human secreted protein encoded by gene 1 clone HGCMD20.	
XX	
KW Human; secreted protein; fusion protein; gene therapy; protein therapy; diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia; developmental abnormality; foetal deficiency; blood; allergy; renal; immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma; inflammation; ischaemic shock; Alzheimer's disease; testostosis; AIDS; cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus; osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion; endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.	
XX	
OS Homo sapiens.	
XX	
FH Key	Location/Qualifiers
FT Misc-difference 97	/label= unknown
FT Misc-difference 119	/label= unknown
FT Misc-difference 240	/label= unknown
XX	
WN WO9839446-A2	

[illegible]

PA	(GETH) GENENTECH INC.
XX	
PI	Chen J, Goddard A, Gurney AL, Pennica D, Wood WL, Yuan J;
DR	WPI; 1999-229533/19.
XX	
DR	N-PSDB; AAX52249.
XX	
PT	New isolated human genes and polypeptides used in, e.g. treatment of
PT	gastrointestinal ulceration
PS	Claim 12: Fig 72; 320pp; English.
XX	
AAV13344-403	represent secreted and transmembrane human proteins.
CC	The CDNA sequences are obtained from cDNA libraries, prepared from
CC	fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
CC	The encoded polypeptides have specific uses based on their homology to
CC	known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
CC	associated with the preservation and maintenance of gastrointestinal
CC	mucosa and the repair of acute and chronic mucosal lesions
CC	(e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
CC	ulceration and congenital microvillus atrophy), skin diseases associated
CC	with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
CC	cancers such as lung squamous cell carcinoma of the vulva and gliomas),
CC	potent effects on cell growth and development, diseases related to
CC	growth or survival of nerve cells including Parkinson's disease,
CC	Alzheimer's disease, ALS, neuropathies or cancer. PRO265 can be used as
CC	for fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
CC	as a target for anti-tumor drugs. PRO533 may be used in the treatment
CC	of Usher Syndrome or Atrophia areata; PRO269 can be used as an
CC	anti-thrombotic agent; PRO287 polypeptides and portions may have
CC	therapeutic applications in wound healing and tissue repair; PRO317 can
CC	be used for treating problems of the kidney, uterus, endometrium, blood
CC	vessels, or related tissue, e.g. in the heart of genital tract.
XX	
SQ	Sequence 467 AA:
Query Match	50.0%; Score 49; DB 20; Length 467;
Best Local Similarity	47.1%; Pred. No. 7.9;
Matches 8; Conservative	1; Mismatches 8; Indels 0; Gaps 0
OY	2 PAMDAESFGAIAHGCL 18 Db 47 pawfdgaqfglfhwgv 63
RESULT	9
AAV94860	
ID	AAV94860 standard; Protein; 467 AA.
XX	
AA	AAV94860;
XX	
DT	12-JUN-2000 (first entry)
XX	
DE	Human protein clone HP02575.
XX	
KW	Human protein; hydrophobic domain; nutritional source; haematopoiesis; cytokine production; cell proliferation; cell differentiation; immune deficiency; infectious disease; autoimmune disorder; asthma; multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis; allergic reaction; osteoporosis; osteoarthritis; periodontal disease; nervous system disorder; Alzheimer's disease; Parkinson's disease; Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury; systemic cytokine damage; tissue differentiation; contraceptive; stroke; coagulation disorder; myocardial infarction; inflammatory condition; septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour; nephritis; therapy.
OS	Homo sapiens.
XX	
XN	WO200005367-A2.
XX	
ND	03-FEB-2000.

PF	22-JUL-1999;	99WO-JP03929.
XX		
PR	24-JUL-1998;	98JP-0208820.
XX	07-AUG-1998;	98JP-0224105.
PR	25-AUG-1998;	98JP-0238116.
PR	09-SEP-1998;	98JP-0254736.
XX	29-SEP-1998;	98JP-0275505.
PA	(SAGA) SAGAMI CHEM RES CENT.	
PA	(PROT-) PROTOGENE INC.	
PI	Kato S, Kimura T;	
DR	WPI; 2000-182694/16.	
XX		
PT	Novel human proteins having hydrophobic domains useful for treating	
PT	osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,	
PT	multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -	
XX		
PS	Claim 1; Page 207-209; 351pp; English.	
XX		
CC	This sequence represents a human protein of the invention, which has	
CC	hydrophobic domains. The DNA sequences can be used as a probe or as a	
CC	genetic marker. The protein can also be used as a marker, and to identify	
CC	potential genetic disorders. The DNA and protein can also be used as	
CC	nutritional sources or supplements. The protein exhibits cytokine, cell	
CC	proliferation, cell differentiation activities and induces production of	
CC	other cytokines in certain cell populations. The protein also exhibits	
CC	immune stimulating or immune suppressing activity. It can be used in the	
CC	treatment of various immune deficiencies and disorders, and to treat	
CC	infectious diseases caused by viral, bacterial, fungal or other	
CC	infections. The protein is also used for treating autoimmune disorders	
CC	such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid	
CC	arthritis. It is also useful in the treatment of allergic reactions and	
CC	conditions such as asthma, and in immune suppression after organ	
CC	transplantation. The protein is useful in regulation of haematopoiesis	
CC	and consequently in the treatment of myeloid or lymphoid cell	
CC	deficiencies. It is also used in compositions for tissue growth or	
CC	regeneration. The protein is also used in the treatment of osteoporosis	
CC	or osteoarthritis and in the treatment of periodontal disease and other	
CC	tooth repair processes. The protein is used in the treatment of nervous	
CC	system disorders such as Alzheimer's disease, Parkinson's disease, and	
CC	Huntington's disease. They are useful for protection or regeneration and	
CC	treatment of lung or liver fibrosis, reperfusion injury in various	
CC	tissues, and conditions resulting from systemic cytokine damage. They are	
CC	also used for promoting or inhibiting tissue differentiation. They are	
CC	also used as contraceptives since they exhibit activin or inhibin related	
CC	activities and as a fertility inducing therapeutic. They are used for	
CC	treating various coagulation disorders and in treatment and prevention of	
CC	conditions resulting from coagulation activities e.g. myocardial	
CC	infarction or stroke. They also acts as receptors, receptor ligands or	
CC	inhibitors or agonists of receptor/ligand interactions. They are used to	
CC	treat inflammatory conditions such as septic shock, sepsis, ischemia	
CC	reperfusion injury, arthritis, nephritis. They can be used to	
CC	prevent tumours.	
XX		
XX		
SQ	Sequence	467 AA;
Query Match	50.0%;	Score 49; DB 21; Length 467;
Best Local Similarity	47.1%;	Pred. No. 7.9;
Matches	8; Conservative	1; Mismatches
		8; Indels 0; Gaps 0;
QY	2 PAMWAESEFGAIHAGL 18	
DB	47 pawdqagqfgifihmgv 63	
RESULT	10	
ID	AAU29031	
AAU29031	standard; Protein; 467 AA.	

AC AAU29031;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human PRO polypeptide sequence #8.
 XX
 KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
 KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200168848-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US06520.
 XX
 PR 01-MAR-2000; 2000WO-US05601.
 PR 02-MAR-2000; 2000WO-US05641.
 PR 03-MAR-2000; 2000US-187202P.
 PR 06-MAR-2000; 2000US-186968P.
 PR 14-MAR-2000; 2000US-189320P.
 PR 14-MAR-2000; 2000US-189328P.
 PR 15-MAR-2000; 2000WO-US06684.
 PR 21-MAR-2000; 2000US-190828P.
 PR 21-MAR-2000; 2000US-191007P.
 PR 21-MAR-2000; 2000US-191048P.
 PR 21-MAR-2000; 2000US-191314P.
 PR 28-MAR-2000; 2000US-192655P.
 PR 29-MAR-2000; 2000US-193032P.
 PR 29-MAR-2000; 2000US-193053P.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 04-APR-2000; 2000US-194449P.
 PR 04-APR-2000; 2000US-194647P.
 PR 11-APR-2000; 2000US-195975P.
 PR 11-APR-2000; 2000US-196000P.
 PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 11-APR-2000; 2000US-196820P.
 PR 18-APR-2000; 2000US-198121P.
 PR 18-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.
 PR 25-APR-2000; 2000US-199654P.
 PR 03-MAY-2000; 2000US-201516P.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 05-JUN-2000; 2000US-209832P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 22-AUG-2000; 2000US-0644848.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32658.
 PR 20-DEC-2000; 2000WO-US34956.
 XX
 PA (GENTH) GENENTECH INC.
 XX
 PI Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI; 2001-602746/68.
 DR N-PSDB; AAS45932.
 XX
 PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 PT to screen for modulators of the compounds -
 XX
 PS Claim 11; Fig 16; 774pp; English.
 XX

CC Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
 CC The PRO polypeptides and their associated nucleic acids can be used to
 CC detect the presence of a tumour in a mammal by comparing the level of
 CC expression of a PRO polypeptide in a test sample of cells from the animal
 CC and a control sample of normal cells, whereby a higher level of
 CC expression in the test sample indicates the presence of a tumour in the
 CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
 CC and rabbits but are preferably human. The polypeptides can be used to
 CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
 CC when contacted with it. A specific polypeptide can be used to stimulate
 CC the proliferation or differentiation of chondrocyte cells. The PRO
 CC proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.
 XX
 SQ Sequence 467 AA;
 XX
 QY 2 PAMDAESFGAIAHGGL 18
 Db 47 pawfdqakfgifhwgv 63
 XX
 RESULT 11
 AAU78342
 ID AAU78342 standard; Protein; 467 AA.
 XX
 AC AAU78342;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human protein SEQ ID NO 1004.
 XX
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX
 PN WO200157190-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US04098.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693325.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
 XX
 DR WPI; 2001-476283/51.
 DR N-PSDB; AAK51475.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX

DR WPI: 1998-230321/20.
DR N-PSDB: AAV21460-61.

XX Isolated cephalosporin esterase gene - obtained from Rhodospiridium
PT toruloides, used to obtain products which can be used for hydrolysis
PT of cephalosporins

XX Claim 13: Fig 5: 38pp; English.

XX This protein comprises Rhodospiridium toruloides ATCC 10657
CC cephalosporium esterase, an enzyme that catalyses the hydrolysis of
CC the 3' acetyl groups of cephalosporins. The enzyme, especially the
CC mature polypeptide, can be purified from cultures of R. toruloides
CC or expressed in claimed host cells (preferably Escherichia coli, R.
CC toruloides, Cephalosporium acremonium or Penicillium chrysogenum)
CC transformed with a vector comprising cephalosporin esterase cDNA or
CC genomic DNA (see AAV21460-61). The isolated polypeptides, even if
CC catalytically inactive, can also be used for the production of
CC antibodies for use in detection and purification.

XX Sequence 572 AA:

Query Match

Best Local Similarity 49.0%; Score 48; DB 19; Length 572;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 2 PAMDAESFGAIAHGGL 18
1 : ||:||: || ||
Db 442 plywtaefgsahkgl 458

Search completed: July 8, 2002, 08:19:12
Job time: 577 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:19:40 ; Search time 21.46 Seconds
(without alignments)
22.764 Million cell updates/sec

Title: US-09-582-296-2

Perfect score: 101

Sequence: 1 APAAVMDAESFGAIHAGL 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	44.6	617	1	US-08-279-700-21
2	43	42.6	330	2	US-08-868-288A-3
3	43	42.6	330	3	US-09-235-373-3
4	43	42.6	330	3	US-09-388-993-3
5	43	42.6	324	4	US-09-230-944-11
6	43	42.6	617	1	US-08-279-700-2
7	43	42.6	617	1	US-08-279-700-4
8	43	42.6	617	1	US-08-279-700-6
9	43	42.6	617	1	US-08-279-700-8
10	43	42.6	617	1	US-08-279-700-10
11	43	42.6	617	1	US-08-279-700-12
12	43	42.6	617	1	US-08-279-700-14
13	43	42.6	617	1	US-08-348-891A-6
14	43	42.6	617	2	US-08-905-817-6
15	43	42.6	617	2	US-09-230-944-10
16	41	40.6	544	2	US-08-932-376A-4
17	41	40.6	572	2	US-08-932-376A-2
18	40	39.6	247	1	US-08-446-083-5
19	39.5	39.1	388	1	US-08-290-448A-80
20	39.5	39.1	388	1	US-08-290-448A-80
21	39.5	39.1	388	1	US-08-175-069A-80
22	39.5	39.1	388	4	US-08-461-939B-80
23	39.5	39.1	388	4	US-08-464-000-80
24	39	38.6	426	6	5221737-2
25	38	37.6	151	2	US-08-485-455D-51
26	38	37.6	151	2	US-08-482-130C-51
27	38	37.6	151	2	US-08-484-211C-51

28	38	37.6	151	3	US-08-906-769-51	Sequence 51, Appl
29	38	37.6	151	3	US-08-906-616-51	Sequence 51, Appl
30	38	37.6	151	4	US-08-817-795-51	Sequence 51, Appl
31	38	37.6	151	4	US-08-485-443B-51	Sequence 51, Appl
32	38	37.6	151	4	US-08-639-075A-51	Sequence 51, Appl
33	38	37.6	151	4	US-09-012-431-51	Sequence 51, Appl
34	38	37.6	151	4	US-09-012-692-51	Sequence 51, Appl
35	38	37.6	151	4	US-08-906-613-51	Sequence 51, Appl
36	38	37.6	151	5	PCT-US95-14442A-51	Sequence 51, Appl
37	38	37.6	344	1	US-08-180-209B-58	Sequence 58, Appl
38	38	37.6	344	4	US-08-474-853-58	Sequence 58, Appl
39	38	37.6	344	5	PCT-US94-02629-58	Sequence 58, Appl
40	38	37.6	496	3	US-08-906-769-113	Sequence 113, Appl
41	38	37.6	496	3	US-08-906-616-113	Sequence 113, Appl
42	38	37.6	496	4	US-08-817-795-113	Sequence 113, Appl
43	38	37.6	496	4	US-08-639-075A-113	Sequence 113, Appl
44	38	37.6	496	4	US-09-012-431-113	Sequence 113, Appl
45	38	37.6	496	4	US-09-012-692-113	Sequence 113, Appl

ALIGNMENTS

RESULT 1
US-08-279-700-21
Sequence 21, Application US/08279700
Patent No. 5578448
GENERAL INFORMATION:
APPLICANT: KOTA, Jennifer S.
APPLICANT: BELINT, William J.
TITLE OF INVENTION: WILD-TYPE MEASLES VIRUS GLYCOPROTEINS:
TITLE OF INVENTION: VACCINE AND DETECTION METHOD THEREFOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/279,700
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/866,033A
FILING DATE: 19920408
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40399/140 NIHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)836-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 617 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
SPRAIN: consensus HA polypeptide
FEATURE:
NAME/KEY: Modified-site
LOCATION: 4
OTHER INFORMATION: /note= "Xaa denotes Gln or His"

1	FEATURE:	NAME/KEY:	Modified-site	
2	LOCATION:	19		
3	OTHER INFORMATION:	/note=	"Xaa denotes Lys or Arg"	
4	FEATURE:	NAME/KEY:	Modified-site	
5	LOCATION:	176		
6	OTHER INFORMATION:	/note=	"Xaa denotes Thr, Val or	
7	OTHER INFORMATION:	Ala"		
8	FEATURE:	NAME/KEY:	Modified-site	
9	LOCATION:	235		
10	OTHER INFORMATION:	/note=	"Xaa denotes Glu or Gly"	
11	FEATURE:	NAME/KEY:	Modified-site	
12	LOCATION:	295		
13	OTHER INFORMATION:	/note=	"Xaa denotes Lys or Arg"	
14	FEATURE:	NAME/KEY:	Modified-site	
15	LOCATION:	303		
16	OTHER INFORMATION:	/note=	"Xaa denotes Glu or Gly"	
17	FEATURE:	NAME/KEY:	Modified-site	
18	LOCATION:	305		
19	OTHER INFORMATION:	/note=	"Xaa represents Ser or Phe"	
20	FEATURE:	NAME/KEY:	Modified-site	
21	LOCATION:	306		
22	OTHER INFORMATION:	/note=	"Xaa denotes Ile or Val"	
23	FEATURE:	NAME/KEY:	Modified-site	
24	LOCATION:	308		
25	OTHER INFORMATION:	/note=	"Xaa denotes Ile or Val"	
26	FEATURE:	NAME/KEY:	Modified-site	
27	LOCATION:	320		
28	OTHER INFORMATION:	/note=	"Xaa denotes Gln or Arg"	
29	FEATURE:	NAME/KEY:	Modified-site	
30	LOCATION:	339		
31	OTHER INFORMATION:	/note=	"Xaa denotes Leu or Phe"	
32	FEATURE:	NAME/KEY:	Modified-site	
33	LOCATION:	348		
34	OTHER INFORMATION:	/note=	"Xaa denotes Arg or Lys"	
35	FEATURE:	NAME/KEY:	Modified-site	
36	LOCATION:	367		
37	OTHER INFORMATION:	/note=	"Xaa denotes Val or Ile"	
38	FEATURE:	NAME/KEY:	Modified-site	
39	LOCATION:	389		
40	OTHER INFORMATION:	/note=	"Xaa denotes Lys or Arg"	
41	FEATURE:	NAME/KEY:	Modified-site	
42	LOCATION:	390		
43	OTHER INFORMATION:	/note=	"Xaa denotes Ile or Asn"	
44	FEATURE:	NAME/KEY:	Modified-site	
45	LOCATION:	446		
46	OTHER INFORMATION:	/note=	"Xaa denotes Ser or Thr"	
47	FEATURE:	NAME/KEY:	Modified-site	
48	LOCATION:	451		
49	OTHER INFORMATION:	/note=	"Xaa denotes Val or Glu"	
50	FEATURE:	NAME/KEY:	Modified-site	
51	LOCATION:	485		
52	OTHER INFORMATION:	/note=	"Xaa denotes Val or Ile"	
53	FEATURE:	NAME/KEY:	Modified-site	
54	LOCATION:	501		
55	OTHER INFORMATION:	/note=	"Xaa denotes Pro or Ser"	

```

1  FEATURE:
2  NAME/KEY: Modified-site
3  LOCATION: 544
4  OTHER INFORMATION: /note="Xaa denotes Ser or Asn"
5  FEATURE:
6  NAME/KEY: Modified-site
7  LOCATION: 546
8  OTHER INFORMATION: /note="Xaa denotes Ser or Gly"
9  FEATURE:
10 NAME/KEY: Modified-site
11 LOCATION: 559
12 OTHER INFORMATION: /note="Xaa denotes Ile or Val"
13 FEATURE:
14 NAME/KEY: Modified-site
15 LOCATION: 560
16 OTHER INFORMATION: /note="Xaa denotes Lys or Arg"
17 FEATURE:
18 NAME/KEY: Modified-site
19 LOCATION: 562
20 OTHER INFORMATION: /note="Xaa denotes Val, Ile or
21 OTHER INFORMATION: Phe"
22 FEATURE:
23 NAME/KEY: Modified-site
24 LOCATION: 593
25 OTHER INFORMATION: /note="Xaa denotes His or Tyr"
26 FEATURE:
27 NAME/KEY: Modified-site
28 LOCATION: 616
29 OTHER INFORMATION: /note="Xaa denotes Arg or Ser"
30 OS-08-279-700-21

```

Query Match	44.6%	Score 45;	DB 1;	Length 617;
Best Local Similarity	50.0%	Pred. No. 9.3;		
Matches	8;	Conservative	3;	Mismatches 5;
				Indels 0;
				Gaps 0.
Qy	5	VMGDAESFGAIAGCL	20	
	1:	1:1111111111		
Db	584	VLADSESGXITRSGM	599	

RESULT 2
 US-08-868-288A-3
 Sequence 3, Application US/08868288A
 Patent No. 5922367
 GENERAL INFORMATION:
 APPLICANT: Au-Young, Janice
 APPLICANT: Lal, Preeti
 APPLICANT: Bandman, Olga
 TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/868,288A
 FILING DATE: June 3, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0309 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 330 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HNT2RAT01
CLONE: 260873
US-08-868-288A-3

Query Match 42.6%; Score 43; DB 2; Length 330;
Best Local Similarity 70.0%; Pred. No. 9.9;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 11 SFGAIAHGGL 20
||||:||||
Db 163 SFGSLGHGGL 172

RESULT 3

US-09-235-373-3
; Sequence 3, Application US/09235373
; Patent No. 6001598
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,373
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE: June 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2RAT01
; CLONE: 260873
US-09-235-373-3

Query Match 42.6%; Score 43; DB 3; Length 330;
Best Local Similarity 70.0%; Pred. No. 9.9;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 11 SFGAIAHGGL 20
||||:||||
Db 163 SFGSLGHGGL 172

RESULT 4

US-09-388-993-3
; Sequence 3, Application US/09388993
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2RAT01
; CLONE: 260873
US-09-388-993-3

Query Match 42.6%; Score 43; DB 3; Length 330;
Best Local Similarity 70.0%; Pred. No. 9.9;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 11 SFGAIAHGGL 20
||||:||||
Db 163 SFGSLGHGGL 172

RESULT 5

US-09-230-944-11
; Sequence 11, Application US/09230944
; Patent No. 6277380
; GENERAL INFORMATION:
; APPLICANT: UEDA, Shigeharu
; APPLICANT: WATANABE, Michiko
; APPLICANT: KAWANISHI, Hitomi
; TITLE OF INVENTION: Measles Virus Mutant Antigen and Gene Coding for the
; TITLE OF INVENTION: Same

MOLECULE TYPE: protein

GENERAL INFORMATION ;

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/279,700
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/866,033A
FILING DATE: 19920408
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40399/140 NIHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)836-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 617 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-279-700-8

Query Match          42.6%; Score 43; DB 1; Length 617;
Best Local Similarity 50.0%; Pseq. No. 20;
Matches      8; Conservative      3; Mismatches      5; Indels      0; Gaps      0

Oy      5 VMGDAESFGAIHAGL 20
       1:|:|:|:|:|:|:|:|:|
Db      584 VLADSESGHITHSGM 599

RESULT 10
US-08-279-700-10
Sequence 10, Application US/08279700
Patent No. 5578448
GENERAL INFORMATION:
APPLICANT: ROTA, Jennifer S.
APPLICANT: BELLINI, William J.
TITLE OF INVENTION: WILD-TYPE MEASLES VIRUS GLYCOPROTEINS;
TITLE OF INVENTION: VACCINE AND DETECTION METHOD THEREFOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/279,700
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/866,033A
FILING DATE: 19920408
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40399/140 NIHD
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)836-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 10:

```


COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/348,891A
FILING DATE: 25-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,400
FILING DATE: 10-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 3-293625
FILING DATE: 14-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J.
REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: KP-7501
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-521-2297
TELEFAX: 703-685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 617 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-348-891A-6

Query Match 42.6%; Score 43; DB 1; Length 617;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 VMGDAESFGAIAHGSL 20
|:|:|:|:|:|:|:|:
Db 584 VLADSESGHITHSGM 599

RESULT 14
US-08-905-817-6
Sequence 6, Application US/08905817
Patent No. 5824777
GENERAL INFORMATION:
APPLICANT: SASAKI, Keiko
APPLICANT: MORI, Takayuki
APPLICANT: MAKINO, Satoshi
TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
NUMBER OF INVENTION: ITS ABSOLUTE IDENTIFICATION
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: YOUNG & THOMPSON
STREET: 745 South 23rd Street
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/905,817
FILING DATE: 04-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/348,891

FILING DATE: 25-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,400
FILING DATE: 10-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 3-293625
FILING DATE: 14-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: PATCH, Andrew J.
REGISTRATION NUMBER: 32,925
REFERENCE/DOCKET NUMBER: KP-7501A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-521-2297
TELEFAX: 703-685-0573
TELEX: 248425 EMBON
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 617 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-905-817-6

Query Match 42.6%; Score 43; DB 2; Length 617;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 VMGDAESFGAIAHGSL 20
|:|:|:|:|:|:|:|:
Db 584 VLADSESGHITHSGM 599

RESULT 15
US-09-230-944-10
Sequence 10, Application US/09230944
Patent No. 6277380
GENERAL INFORMATION:
APPLICANT: UEDA, Shigeharu
APPLICANT: WATANABE, Michiko
APPLICANT: KAWANISHI, Hitomi
TITLE OF INVENTION: Measles Virus Mutant Antigen and Gene Coding for the
FILE REFERENCE: 0216-0407P
CURRENT APPLICATION NUMBER: US/09/230,944
EARLIER FILING DATE: 1999-02-04
EARLIER APPLICATION NUMBER: PCT/JP98/02481
EARLIER FILING DATE: 1998-06-04
EARLIER APPLICATION NUMBER: 9-184285 JAPAN
EARLIER FILING DATE: 1997-06-04
NUMBER OF SEQ ID NOS: 32
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 617
TYPE: PRT
ORGANISM: Measles virus
FEATURE:
OTHER INFORMATION: any n or xaa = Unknown
US-09-230-944-10

Query Match 42.6%; Score 43; DB 4; Length 617;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 5 VMGDAESFGAIAHGSL 20
|:|:|:|:|:|:|:|:
Db 584 VLADSESGHITHSGM 599

Search completed: July 8, 2002, 08:19:41
Job time: 586 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:20:15 ; Search time 26.89 Seconds

(without alignments)
71.468 Million cell updates/sec

Title: US-09-582-296-2

Perfect score: 101

Sequence: 1 APAYVMGDAESFGAIAHGGL 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_71:*
2: pir1:*
3: pir2:*
4: pir3:*
5: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	47.5	749	2	S75331
2	48	47.5	750	1	H70628
3	48	47.5	763	2	T44735
4	48	47.5	767	2	H86946
5	46	45.5	121	2	F83769
6	45.5	45.0	1513	2	A70982
7	45	44.6	583	2	G81412
8	45	44.6	746	2	H84301
9	44.5	44.1	469	2	A81836
10	44.5	44.1	469	2	G81128
11	44.5	44.1	474	2	G83255
12	44.5	44.1	474	2	H82564
13	44	43.6	242	2	A11883
14	44	43.6	498	2	T37527
15	43.5	43.1	469	2	AD2909
16	43.5	43.1	469	2	AH3271
17	43.5	43.1	475	2	C97684
18	43.5	43.1	479	2	C87273
19	43.5	43.1	552	2	T07932
20	43.5	43.1	2946	2	T00867
21	43	42.6	266	2	AH1055
22	43	42.6	387	2	AB2977
23	43	42.6	395	2	AB2977
24	43	42.6	422	2	A98306
25	43	42.6	617	1	HNNZHA
26	43	42.6	617	1	HNNZED
27	43	42.6	617	1	F48556
28	42.5	42.1	205	2	F42774
29	42	41.6	242	2	G69128

30	42	41.6	329	1	G69210	conserved hypothet
31	42	41.6	374	2	F87280	esterase A [import
32	42	41.6	618	2	A70989	hypothetical glyci
33	42	41.6	898	2	JH0440	RAD54 protein - ye
34	42	41.6	1126	2	A96032	probable two-compo
35	41.5	41.1	354	2	T46683	probable lipote p
36	41.5	41.1	358	2	H90047	hypothetical prote
37	41.5	41.1	388	2	T28693	hypothetical prote
38	41	40.6	61	2	F95014	hypothetical prote
39	41	40.6	61	2	H97887	hypothetical prote
40	41	40.6	221	2	S47690	hypothetical 25.3k
41	41	40.6	221	2	H91168	hypothetical prote
42	41	40.6	221	2	H86014	hypothetical prote
43	41	40.6	223	2	E87590	DNA-binding respon
44	41	40.6	270	2	C87568	transcription regu
45	41	40.6	323	2	A70404	acetyl-CoA carboxy

ALIGNMENTS

RESULT 1
S75331
penicillin-binding protein 1b mrcB - *Synechocystis* sp. (strain PCC 6803)
N:Alternate names: protein slr1710
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S75331
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O., K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocys*
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S75331
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-749 <KAN>
A:Cross-references: EMBL:D90904; GB:AB001339; NID:g1652225; PIDD:BA17245.1; PTD:g165
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: mrcB
C:Superfamily: penicillin-binding protein

Query Match 47.5%; Score 48; DB 2; Length 749;
Best Local Similarity 37.0%; Pred. No. 13;
Matches 10; Conservative 7; Mismatches 2; Indels 8; Gaps 1;

OY 2 APAYVMGDAE-----SFGAIAHGGL 20
Db 592 PGLVIGSEARVLEMTGAYGIANGV 618

RESULT 2
H70628
probable pkng protein - *Mycobacterium tuberculosis* (strain H37RV)
C:Species: *Mycobacterium tuberculosis*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: H70628
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete geno
A:Reference number: A70500; MUID:98255987
A:Accession: H70628
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-750 <COL>
A:Cross-references: GB:Z84724; GB:AL123456; NID:g3261708; PIDD:CA060580.1; PTD:g18176

A:Experimental source: strain H37RV
 C:Gene: pknG
 C:Superfamily: Mycobacterium tuberculosis probable pknG protein: protein kinase homology
 F:149-393/Domain: protein kinase homology <KIN>

Query Match 47.5%; Score 48; DB 1; Length 750;
 Best Local Similarity 58.8%; Pred. No. 13;
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 VVMGDAESFGAIAHGCL 20
 : | | | | | | | | | |
 Db 146 IVAGQYEVKGCIAHGCL 162

RESULT 3
 T44735
 serine/threonine-specific protein kinase homolog [imported] - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 04-Mar-2000
 C:Accession: T44735
 R:James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 Submitted to the EMBL Data Library, January 1998
 A:Reference number: Z22831
 A:Accession: T44735
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-763 <JAM>
 A:Cross-references: EMBL:AL035159; PIDN:CAA22703.1
 A:Experimental source: cosmid B1450
 C:Genetics:
 A:Note: MLCB1450.19c
 C:Superfamily: Mycobacterium tuberculosis probable pknG protein: protein kinase homology

Query Match 47.5%; Score 48; DB 2; Length 763;
 Best Local Similarity 58.8%; Pred. No. 13;
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 VVMGDAESFGAIAHGCL 20
 : | | | | | | | | | |
 Db 155 IVAGQYEVKGCIAHGCL 171

RESULT 4
 H86946
 Probable serine-threonine protein kinase [imported] - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
 C:Accession: H86946
 R:Cole, S.T.; Eiglsamer, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
 R.; Davies, R.M.; Devlin, K.; Duchoy, S.; Felwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
 eam, M.A.; Rutherford, K.M.
 Nature 409, 1007-1011, 2001
 A:Authors: Rutherford, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
 A:Title: Massive gene decay in the leprosy bacillus.
 A:Reference number: A86909; MUID:21128732; PMID:11234002
 A:Accession: H86946
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-767 <STO>
 A:Cross-references: GB:AL450380; NID:g13092623; PIDN:CAC29812.1; GSPDB:GN00147
 C:Genetics:
 A:Gene: pknG
 C:Superfamily: Mycobacterium tuberculosis probable pknG protein: protein kinase homology

Query Match 47.5%; Score 48; DB 2; Length 767;
 Best Local Similarity 58.8%; Pred. No. 14;
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 VVMGDAESFGAIAHGCL 20

Db 159 IVAGQYEVKGCIAHGCL 175
 : | | | | | | | | | |

RESULT 5
 F83769
 hypothetical protein BH0958 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: F83769
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: F83769
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-121 <STO>
 A:Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04677.1; GSPDB:G
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH0958

Query Match 45.5%; Score 46; DB 2; Length 121;
 Best Local Similarity 56.2%; Pred. No. 3.7;
 Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 5 VVMGDAESFGAIAHGCL 20
 : | | | | | | | | | |
 Db 39 MMGDLEVFISHVAGL 54

RESULT 6
 A70982
 Probable ATP-dependent helicase lhr - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: A70982
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
 A:Reference number: A70500; MUID:98295587
 A:Accession: A70982
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1513 <COL>
 A:Cross-references: GB:U22771; GB:AL123456; NID:g3242259; PIDN:CAB07060.1; PID: e30655
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: RV3296; lhr

Query Match 45.0%; Score 45.5; DB 2; Length 1513;
 Best Local Similarity 45.8%; Pred. No. 73;
 Matches 11; Conservative 3; Mismatches 5; Indels 5; Gaps 1;

QY 1 APAVVMGDAESFG-----AIAHGG 19
 : | | | | | | | | | |
 Db 316 APAHMGSSQGTGAPPVLAARAHG 339

RESULT 7
 G81412
 aspartate-tRNA ligase (PC 6.1.1.12) Cj0640c [imported] - Campylobacter jejuni (strai
 C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Jun-2000
 C:Accession: G81412
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chl
 C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba

Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyf
A:Reference number: AB1250; MUID:20150912
A:Accession: G81412
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-563 <PAR>
A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB75276.1; PID:g696810
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: asps; Cj0640c
C:Superfamily: lysine--tRNA ligase
C:Keywords: ligase

Query Match 44.6%; Score 45; DB 2; Length 583;
Best Local Similarity 69.2%; Pred. No. 30;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 8 DAEFGAIAHGCL 20
||| ||| |||
DB 514 DALSFCAPHGCI 526

RESULT 8
H84301
hypothetical protein Vng1476c [imported] - *Halobacterium* sp. NRC-1
C:Species: *Halobacterium* sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84301
R:Ng, W.V.; Kennedy, S.P.; Mahatras, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
; Leitman, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alm, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of *Halobacterium* species NRC-1.
A:Reference number: AB4160; MUID:20504483
A:Accession: H84301
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-746 <STO>
A:Cross-references: GB:AE004437; NID:g10580974; PIDN:AG19780.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG1476C

Query Match 44.6%; Score 45; DB 2; Length 746;
Best Local Similarity 50.0%; Pred. No. 40;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 AYYMGDAESFGAIAHGCL 20
||| ||| |||
DB 220 AFEVGYITFGAIGYGI 237

RESULT 9
AB1836
probable 3-isopropylmalate dehydratase (EC 4.2.1.33) large chain NMA1450 [imported] - *Ne*
C:Species: *Neisseria meningitidis*
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: AB1836
R:Parhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churche, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A:Reference number: AB1775; MUID:2022556
A:Accession: AB1836
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-469 <PAR>
A:Cross-references: GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB84686.1; PID:g738010
A:Experimental source: serogroup A, strain Z2491
C:Genetics:

A:Gene: leuc; NMA1450
C:Superfamily: aconitate hydratase
C:Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 44.1%; Score 44.5; DB 2; Length 469;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 10; Conservative 3; Mismatches 2; Indels 5; Gaps 1;

QY 4 VYMGDAE-----SFGAIAHG 18
||| ||| |||
DB 126 VVCGDSHTSTHGAFGALAHG 145

RESULT 10
G81128
3-isopropylmalate dehydratase, large chain NMB1036 [imported] - *Neisseria meningitidis*
C:Species: *Neisseria meningitidis*
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: G81128
R:Teitelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
; Li, H.; Qin, H.; Yatchewyan, J.; Gill, J.; Scariato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.
A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755
A:Accession: G81128
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-469 <TER>
A:Cross-references: GB:AE002454; GB:AE002098; NID:g7226273; PIDN:AAF41435.1; PID:g722
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1036
C:Superfamily: aconitate hydratase

Query Match 44.1%; Score 44.5; DB 2; Length 469;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 10; Conservative 3; Mismatches 2; Indels 5; Gaps 1;

QY 4 VYMGDAE-----SFGAIAHG 18
||| ||| |||
DB 126 VVCGDSHTSTHGAFGALAHG 145

RESULT 11
G83255
3-isopropylmalate dehydratase large subunit PA3121 [imported] - *Pseudomonas aeruginos*
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83255
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: G83255
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <STO>
A:Cross-references: GB:AE004736; GB:AE004091; NID:g9949227; PIDN:AA06509.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: leuc; PA3121
C:Superfamily: aconitate hydratase

Query Match 44.1%; Score 44.5; DB 2; Length 474;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 10; Conservative 3; Mismatches 2; Indels 5; Gaps 1;

Oy 4 VWMGDAB-----SFGAIAHG 18
|||:|||||
Db 130 VVCGDSHTSTHGAFGALAHG 149

RESULT 12

H82564
3-isopropylmalate dehydratase large subunit XF2375 [imported] - Xylella fastidiosa (strain
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: H82564
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: H82564
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <SIM>
A:Cross-references: GB:AE004047; GB:AE003849; NID:g9107548; PIDN:AAE85174.1; GSPDB:GN001
A:Experimental source: strain 945C
R:Simposon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briomes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carreir
as-Neto, E.; Docena, C.; El-Dorri, H.; Facinca, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurama, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Savasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF2375
C:Superfamily: aconitate hydratase

Query Match 44.1%; Score 44.5; DB 2; Length 474;
Best Local Similarity 50.0%; Pred. No. 29;
Matches 10; Conservative 3; Mismatches 2; Indels 5; Gaps 1;
Oy 4 VWMGDAB-----SFGAIAHG 18
|||:|||||
Db 130 VVCGDSHTSTHGAFGALAHG 149

RESULT 13

A11883
hypothetical protein alr0618 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: A11883
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuch
Nakazaki, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genome Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: A11883
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-242 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAE72576.1; PID:g17129964; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0618

Query Match 43.6%; Score 44; DB 2; Length 242;
Best Local Similarity 50.0%; Pred. No. 17;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Oy 4 VWMGDAB-----SFGAIAHG 19
|||:|||||
Db 215 IILDAEKLNAIAYCG 230

RESULT 14

T37527
probable allantoinase transporter - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: T37527
R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z21720
A:Accession: T37527
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-498 <CON>
A:Cross-references: EMBL:Z68166; PIDN:CAA92319.1; GSPDB:GN00066; SPDB:SPAC11D3.18c
A:Experimental source: strain 972h-; cosmid c11D3
C:Genetics:
A:Gene: SPDB:SPAC11D3.18c
A:Map position: 1
C:Superfamily: conserved probable membrane protein YCR028C

Query Match 43.6%; Score 44; DB 2; Length 498;
Best Local Similarity 40.0%; Pred. No. 37;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Oy 1 APVWMDAB-----SFGAIAHGL 20
|||:|||||
Db 438 APKYILGNAPFLGCVIGL 457

RESULT 15

AD2909
3-isopropylmalate dehydratase, large subunit [imported] - Agrobacterium tumefaciens (C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AD2909
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; MCCI
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AD2909
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-469 <KUR>
A:Cross-references: GB:AE008688; PIDN:ALA43690.1; PID:g17741217; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: leuc
A:Map position: circular chromosome
C:Superfamily: aconitate hydratase

Query Match 43.1%; Score 43.5; DB 2; Length 469;
Best Local Similarity 45.0%; Pred. No. 42;
Matches 9; Conservative 4; Mismatches 2; Indels 5; Gaps 1;

Oy 4 VWMGDAB-----SFGAIAHG 18
|||:|||||
Db 127 IVCGDSHTSTHGAFGALAHG 146

* Mon Jul 8 09:33:17 2002

us-09-582-296-2.rpt

Page 5

Search completed: July 8, 2002, 08:20:16
Job time: 576 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:21:22 ; Search time 13.48 seconds
(without alignments)
57.447 Million cell updates/sec

Title: US-09-582-296-2

Perfect score: 101
Sequence: 1 APAYVMGDAESFGAIAHGSL 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SWISSProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	47.5	750	1 PKNG_MYCTU	P66256 mycobacteri
2	48	47.5	763	1 PKNG_MYCLE	P57993 mycobacteri
3	44.5	44.1	323	1 AOP4_BOVIN	077750 bos taurus
4	44.5	44.1	323	1 AOP4_HUMAN	P55087 homo sapien
5	44	43.6	498	1 YAOI_SCHPO	Q10097 schizosach
6	43	42.6	326	1 DJB6_HUMAN	Q75150 homo sapien
7	43	42.6	617	1 HEMA_MEASA	P35971 measles vir
8	43	42.6	617	1 HEMA_MEASH	P08362 measles vir
9	43	42.6	617	1 HEMA_MEASH	P66030 measles vir
10	43	42.6	617	1 HEMA_MEASH	Q13607 measles vir
11	42	41.6	317	1 Q2FL_HUMAN	Q26918 homo sapien
12	42	41.6	329	1 Y830_MERTH	Q68918 methanobact
13	42	41.6	898	1 RAS4_YEAST	P32863 saccharomyc
14	41	40.6	185	1 APT_CORGL	P87330 corynebacte
15	41	40.6	221	1 YHHQ_ECOLI	P37619 escherichia
16	41	40.6	323	1 ACCA_AQUAE	Q67260 aquifex aeo
17	41	40.6	465	1 LUDH_CHLTR	Q84561 chlamydia t
18	40.5	40.1	357	1 YNS7_YEAST	P53873 saccharomyc
19	40.5	40.1	424	1 SYH_BACSU	Q32039 bacillus su
20	40.5	40.1	968	1 Y682_MERVA	P68095 methanococ
21	40	39.6	282	1 YOR3_MERVA	P68080 methanococ
22	40	39.6	317	1 Q2F2_HUMAN	Q95006 homo sapien
23	40	39.6	591	1 SYD_PSEAE	Q51452 pseudomonas
24	39.5	39.1	314	1 VANB_PSEAS	P12580 pseudomonas
25	39.5	39.1	448	1 GCSI_BACSU	P54376 bacillus su
26	39.5	39.1	661	1 Y182_SYNY3	Q55774 synechocyst
27	39	38.6	151	1 AERC_AERSO	P09165 aeromonas s
28	39	38.6	151	1 RS9_MYCTU	P06259 mycobacteri
29	39	38.6	242	1 DJB6_MOUSE	Q54946 mus musculu
30	39	38.6	291	1 TSPA_MICLU	Q66126 mtococcus
31	39	38.6	328	1 NTR3_AZOBR	P45672 azospirillum
32	39	38.6	342	1 EFTS_LACIA	Q9cd55 lactococcus
33	39	38.6	345	1 EFTS_STRPN	P80715 streptococc

34	39	38.6	426	1 GABT_ECOLI	P22256 escherichia
35	39	38.6	453	1 VG_DROME	Q26366 pseudophila
36	39	38.6	480	1 CALB_PSESP	O86447 pseudomonas
37	39	38.6	481	1 GLNA_HELPY	P94845 helicobacte
38	39	38.6	484	1 US15_HCMVA	P09718 human cytom
39	39	38.6	528	1 GD_DROME	O62589 drosophila
40	39	38.6	587	1 RECN_MYCTU	O33197 mycobacteri
41	39	38.6	1051	1 UBA1_WHEAT	P20973 triticum ae
42	39	38.6	1051	1 UBA2_WHEAT	P31251 triticum ae
43	39	38.6	1091	1 DIA_DROME	P48608 drosophila
44	39	38.6	1197	1 EVGS_ECO57	P58402 escherichia
45	39	38.6	1197	1 EVGS_ECOLI	P30855 escherichia

ALIGNMENTS

RESULT	ID	PKNG_MYCTU	STANDARD:	PRT:	750 AA.
1	PKNG_MYCTU				
AC	P66256				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Probable serine/threonine-protein kinase pknG (EC 2.7.1.-).				
GN	PKNG OR RV0410C OR MT0423 OR MTCY22G10.06C.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;				
OC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37R;				
RX	MDL=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eigler K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,				
RA	Hornby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,				
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;				
RT	Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence."				
RL	Nature 393:537-544(1998).				
[2]					
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CDC 1551 / Oshkosh;				
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,				
RA	Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,				
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,				
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,				
RT	Bisai W.;				
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and				
RT	laboratory strains."				
RT	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.				
RT	-1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.				
CC					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/				
CC	or send an email to license@isb-sdb.ch).				
CC					
DR	EMBL; 284724; CAB06580.1; -				
DR	EMBL; AE006946; AAK4647.1; ALT_INT.				
DR	TIGR; MT0423; -				
DR	Tuberculint; RV0410c; -				
DR	InterPro; IPR000719; Euk_pkinase.				
DR	InterPro; IPR002290; Ser_thr_pkinase.				
DR	Pfam; PF00069; pkinase; 1.				

DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
 KM ATP-binding; Complete proteome.
 FT DOMAIN 151 396
 FT NP_BIND 157 165 ATP (BY SIMILARITY).
 FT BINDING 181 181 ATP (BY SIMILARITY).
 FT ACT_SITE 276 276 BY SIMILARITY.
 SQ SEQUENCE 750 AA; 81577 MW; CFC6E54DBE19569F CRC64;

Query Match 47.5%; Score 48; DB 1; Length 750;
 Best Local Similarity 58.8%; Pred. No. 6.6;
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 VYMGDAEFGAIGHGL 20
 : | | | | | | | | | |
 Db 146 IVAGQYEVKGCIAHGGL 162

RESULT 2
 PKNG_MYCLE STANDARD; PRT; 763 AA.
 AC P57993; Q9ZBL8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Probable serine/threonine-protein kinase pknd (EC 2.7.1.-).
 GN PKNG OR M0304 OR M0304.19C.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies K.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagsels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrall B.G.,
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL035159; CA22703.1; -
 DR EMBL: AL583918; CAC29812.1; ALU_INIT.
 DR Leptoma; ML0304;
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002290; Ser_thr_pkinase.
 DR InterPro: IPR001245; Tyr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR SMART: SM00220; S_TKC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 DR PROSITE: PS0011; PROTEIN_KINASE_DOM; 1.
 KW Hypothetical protein; Transferase; Serine/threonine-protein kinase;
 KM ATP-binding; Complete proteome.

FT DOMAIN 160 406 PROTEIN_KINASE.
 FT NP_BIND 166 174 ATP (BY SIMILARITY).
 FT BINDING 190 190 ATP (BY SIMILARITY).
 FT ACT_SITE 289 289 BY SIMILARITY.
 SQ SEQUENCE 763 AA; 83469 MW; 942725EB49C1199 CRC64;

Query Match 47.5%; Score 48; DB 1; Length 763;
 Best Local Similarity 58.8%; Pred. No. 6.7;
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 VYMGDAEFGAIGHGL 20
 : | | | | | | | | | |
 Db 155 IVAGQYEVKGCIAHGGL 171

RESULT 3
 AQP4_BOVIN STANDARD; PRT; 323 AA.
 AC O77750;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Aquaporin 4 (WCH4) (Mercurial-insensitive water channel) (MIMC).
 GN AQP4 OR AQP-4.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC STRAIN=Holstein; TISSUE=Brain;
 RX PubMed=10673041;
 RA Sobue K., Yamamoto N., Yoneda K., Fujita K., Miura Y., Asai K.,
 RA Tsuda T., Katsuya H., Kato T.;
 RT "Molecular cloning of two bovine aquaporin-4 cDNA isoforms and their
 RT expression in brain endothelial cells."
 RL Biochim. Biophys. Acta 1489:393-398(1999).
 CC -1- FUNCTION: FORMS A WATER-SPECIFIC CHANNEL. OSMORECEPTOR WHICH
 CC REGULATES BODY WATER BALANCE AND MEDIATES WATER FLOW WITHIN THE
 CC CENTRAL NERVOUS SYSTEM.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1/A and 2/B (shown here); are
 CC produced by alternative splicing.
 CC -1- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB015947; BAA36505.2; -
 DR EMBL: AB012950; BAA33583.1; -
 DR EMBL: AB028642; BAA89291.1; -
 DR HSSP: P29972; IROY.
 DR InterPro: IPR00425; MIP.
 DR Pfam: PF00230; MIP; 1.
 DR PRINTS: PRO0783; MINTRINSICP.
 DR PROSITE: PS00221; MIP; 1.
 KM Transport; Transmembrane; Alternative splicing.
 FT DOMAIN 1 36
 FT TRANSMEM 1 36 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 37 57 POTENTIAL.
 FT TRANSMEM 58 64 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 65 85 POTENTIAL.
 FT TRANSMEM 86 115 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 116 136 POTENTIAL.
 FT TRANSMEM 137 135 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 156 176 POTENTIAL.
 FT TRANSMEM 177 184 CYTOPLASMIC (POTENTIAL).

```

FT TRANSMEM 185 205 POTENTIAL.
FT DOMAIN 206 231 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 232 252 POTENTIAL.
FT DOMAIN 253 323 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 22 MISSING (IN ISOFORM 1).
SQ SEQUENCE 323 AA: 34616 MW: EBE14507B6474D5F CXC64;

Query Match 44.1%; Score 44.5; DB 1: Length 323;
Best Local Similarity 39.3%; Pred. No. 10;
Matches 11; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

OY 2 PAVVMGDAES-----FGAIAHG L 20
Db 220 PAVINGMNMHWYVWGPIIGAVLGG L 247
|||:::|: ||: |||
|||:::|: ||: |||

RESULT 4
AOP4_HUMAN STANDARD: PRT: 323 AA.
ID AOP4_HUMAN
AC P55087; P78564;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aquaporin 4 (wCh4) (Mercurial-insensitive water channel) (MITC).
OS AOP4.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Fetal brain;
RC MEDLINE=96032721; PubMed=7559426;
RA Yang B., Ma T., Verkman A.S.;
RT "cDNA cloning, gene organization, and chromosomal localization of a
RT human mercurial insensitive water channel. Evidence for distinct
RL J. Biol. Chem. 270:22907-22913(1995).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=96176324; PubMed=8601457;
RX Misaka T., Abe K., Iwabuchi K., Kusakabe Y., Ichinose M., Miki K.,
RA Emori Y., Arai S.;
RT "A water channel closely related to rat brain aquaporin 4 is
RT expressed in acid- and pepsinogen-secretory cells of human stomach.";
RL FEBS Lett. 361:208-212(1996).
[3]
RN SEQUENCE FROM N.A.
RP MEDLINE=97008105; Pubmed=8855281;
RA Lu M., Lee M.D., Smith B.L., Jung J.S., Agre P., Verdijk M.A.J.,
RA Merx G., Rijss J.P.L., Deen P.M.T.;
RT "The human AQP4 gene: definition of the locus encoding two water
RT channel polypeptides in brain.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:10908-10912(1996).
CC -!- FUNCTION: FORMS A WATER-SPECIFIC CHANNEL, OSMORECEPTOR WHICH
CC REGULATES BODY WATER BALANCE AND MEDIATES WATER FLOW WITHIN THE
CC CENTRAL NERVOUS SYSTEM.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 AND 2 (SHOWN HERE); ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: BRAIN - MUSCLE >> HEART, KIDNEY, LUNG, AND
CC TRACHEA.
CC -!- SIMILARITY: BELONGS TO THE MIP/AQUAPORIN FAMILY (TC 1.A.8).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
```

```

CC      or send an email to license@lsb-sib.ch).
-----
CC      EMBL; U34846; AAC52112.1; ALT_INIT.
CC      DR      EMBL; U34845; AAC50284.1; ALT_INIT.
CC      DR      EMBL; D63412; BAA09715.1; -.
CC      DR      EMBL; U63622; AAB26957.1; -.
CC      DR      EMBL; U63623; AAB26958.1; -.
CC      DR      HSEP; P29972; IFOY.
CC      DR      MIM; 600308; -.
CC      DR      InterPro; IPR000425; MTP.
CC      DR      Pfam; PF00230; MTP; 1.
CC      DR      PRINTS; PR00783; MINTIRNSICP.
CC      DR      PROSITE; PS00221; MTP; 1.
CC      KM      Transport; Transmembrane; Alternative splicing.
CC      FT      DOMAIN 1 36 CYTOPLASMIC (POTENTIAL).
CC      FT      TRANSMEM 37 57 POTENTIAL.
CC      FT      DOMAIN 58 64 EXTRACELLULAR (POTENTIAL).
CC      FT      TRANSMEM 65 85 POTENTIAL.
CC      FT      DOMAIN 86 115 CYTOPLASMIC (POTENTIAL).
CC      FT      TRANSMEM 116 136 POTENTIAL.
CC      FT      DOMAIN 137 155 EXTRACELLULAR (POTENTIAL).
CC      FT      TRANSMEM 156 176 POTENTIAL.
CC      FT      DOMAIN 177 184 CYTOPLASMIC (POTENTIAL).
CC      FT      TRANSMEM 185 205 POTENTIAL.
CC      FT      DOMAIN 206 231 EXTRACELLULAR (POTENTIAL).
CC      FT      TRANSMEM 232 252 POTENTIAL.
CC      FT      DOMAIN 253 323 CYTOPLASMIC (POTENTIAL).
CC      FT      CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      CARBOHYD 206 206 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC      FT      VARSPPLIC 1 22 MISSING (IN ISOFORM 1).
CC      FT      CONFLICT 246 246 G -> A (IN REF. 1).
CC      FT      CONFLICT 287 288 VE -> AK (IN REF. 1).
CC      FT      CONFLICT 296 296 P -> L (IN REF. 1).
CC      SQ      SEQUENCE 323 AA; 34829 MM; 1A1600CFDDC11052 CRC64;
CC
CC      Query Match 44.1%; Score 44.5; DB 1; Length 323;
CC      Best Local Similarity 39.3%; Pred. No. 10;
CC      Matches 11; Conservative 4; Mismatches 4; Indels 9; Gaps 1;
CC
CC      QY      2 PAVVMGDAES-----FGAHAGGL 20
CC      DB      220 PAVINGNWNEMWYVGPILGAVLAGGL 247
CC
CC      RESULT 5
CC      YAOI_SCHPO
CC      ID YAOI_SCHPO STANDARD: PRT; 498 AA.
CC      AC Q10097.
CC      DT 01-FEB-1996 (Rel. 33, Created)
CC      DT 01-FEB-1996 (Rel. 33, Last sequence update)
CC      DT 15-JUL-1998 (Rel. 36, Last annotation update)
CC      DE Putative transporter C11D3.18C.
CC      GN SPAC11D3.18C.
CC      OS Schizosaccharomyces pombe (Fission yeast).
CC      OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
CC      OC Schizosaccharomycetales; Schizosaccharomycetaceae;
CC      OC Schizosaccharomycetes.
CC      OX NCBI_TaxID=4896;
CC      RN [1]
CC      RP SEQUENCE FROM N.A.
CC      RC STRAIN=972;
CC      RL Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
CC      Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
CC      CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC      CC -1- SIMILARITY: BELONGS TO THE ALANTOATE PERMEASE FAMILY.
CC      CC -----
CC      CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      CC the European Bioinformatics Institute. There are no restrictions on its
CC      CC use by non-profit institutions as long as its content is in no way
CC      CC modified and this statement is not removed. Usage by and for commercial
CC      CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

or send an email to license@isb-sib.ch.

CC EMBL: 268166; CA92319.1; -
 DR Hypothetical protein: Transmembrane; Transport.
 KW TRANSMEM 54 74 POTENTIAL.
 FT TRANSMEM 100 120 POTENTIAL.
 FT TRANSMEM 128 148 POTENTIAL.
 FT TRANSMEM 150 170 POTENTIAL.
 FT TRANSMEM 188 208 POTENTIAL.
 FT TRANSMEM 221 241 POTENTIAL.
 FT TRANSMEM 302 322 POTENTIAL.
 FT TRANSMEM 326 346 POTENTIAL.
 FT TRANSMEM 353 373 POTENTIAL.
 FT TRANSMEM 381 401 POTENTIAL.
 FT TRANSMEM 446 466 POTENTIAL.
 SO SEQUENCE 498 AA; 55351 MW; B1409306517DBE49 CRC64;

Query Match 43.6%; Score 44; DB 1; Length 498;
 Best Local Similarity 40.0%; Pred. No. 19;
 Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

Qy 1 APAYVMDAESFGAIHAGL 20
 11 : 11 : 11 :
 Db 438 APXYILGNATFLGCVIGGL 457

RESULT 6

DJB6_HUMAN STANDARD; PRT; 326 AA.
 AC 075190; Q95806; Q90IK6;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DnaJ homolog subfamily B member 6 (Heat shock protein J2) (HSJ-2)
 DE (MSJ-1) (HHDJ1) (MKJ).
 GN DNAB6 OR HSJ2 OR MSJ1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RC TISSUE=Testis;
 RA Saito T., Seki N.;
 RT "A new member of human dnaJ-related gene family";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RA Zhang W., Wan T., Yuan Z., Cao X.;
 RT "HSJ2, a novel human homologue of the bacterial heat-shock protein
 DnaJ";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RC TISSUE=Testis;
 RX MEDLINE=99115663; Pubmed=9915854;
 RA Pel L.;
 RT "Pituitary tumor-transforming gene protein associates with ribosomal
 protein S10 and a novel human homologue of DnaJ in testicular cells";
 RL J. Biol. Chem. 274:3151-3158(1999).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORMS AND B).
 RA Hanai R.;
 RT "DnaJ homologue HSJ2b";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RA Zhang J.S., Nelson M., Wang L., Smith D.L.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RC TISSUE=Kidney;

RX MEDLINE=21154917; Pubmed=11230166;
 RA Wernann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
 RA Ansgorge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,
 RA Lauber J., Duesthoeft A., Beyer A., Koehrer K., Strack N.,
 RA Mewes H.-W., Oltmerhoeft B., Obermaier B., Tampe J., Heubner D.,
 RA Wambutt R., Korn B., Klein M., Poustka A.;
 RT "Towards a catalog of human genes and proteins: sequencing and
 RT analysis of 500 novel complete protein coding human cDNAs";
 RL Genome Res. 11:422-435(2001).
 RN [7]
 RP SEQUENCE FROM N.A. (ISOFORMS AND B).
 RC TISSUE=Skin, and Placenta;
 RA Strausberg R.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: INTERACTS WITH PTG.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; a (shown here) and b; are
 CC produced by alternative splicing.
 CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to a
 CC frameshift in position 197.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AB014888; BAA32209.1; -
 DR EMBL: AF075601; AAD43194.1; -
 DR EMBL: AF080569; AAD16010.1; ALT_FRAME.
 DR EMBL: AB015798; BAA88769.1; -
 DR EMBL: AB015799; BAA88770.1; -
 DR EMBL: AF060703; AAF21257.1; -
 DR EMBL: AL136707; CAB66642.1; -
 DR EMBL: BC000177; AAH00177.1; -
 DR EMBL: BC002446; AAH02446.1; -
 DR HSSP: P25685; 1HDJ.
 DR InterPro: IPR001623; DnaJ_N.
 DR Pfam: PF00226; DnaJ; 1.
 DR SMART: SM00271; DnaJ; 1.
 DR PROSITE: PS00636; DnaJ_1; 1.
 DR PROSITE: PS50076; DnaJ_2; 1.
 KW Chapterone; Alternative splicing.
 FT DOMAIN 1 69 GLY/PHE-RICH.
 FT DOMAIN 83 172 VADDDALAE -> KEQLLRIDNK (IN ISOFORM B).
 FT VARSPLIC 232 241 MISSING (IN ISOFORM B).
 FT VARSPLIC 242 326
 SO SEQUENCE 326 AA; 36087 MW; ECF1628BFA524F3 CRC64;

Query Match 42.6%; Score 43; DB 1; Length 326;
 Best Local Similarity 70.0%; Pred. No. 18;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 11 SFGAIHAGL 20
 11 : 11 : 11 :
 Db 163 SFGSLGHGGL 172

RESULT 7

HEMA_MEASA STANDARD; PRT; 617 AA.
 AC P35971;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin-neuraminidase (EC 3.2.1.18).
 GN HN.
 OS Measles virus (strain Aik-C) (Subacute sclerosing panencephalitis
 OS virus).
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;

CC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
RX NCBI_TaxID=36408;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=93227570; PubMed=8470368;
RT Mori T., Sasaki K., Hashimoto H., Makino S.;
RT "Molecular cloning and complete nucleotide sequence of genomic RNA of
RT the A/C strain of attenuated measles virus.";
RT Virus Genes 7:67-81(1993).
CC
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS
CC TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE
CC ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING
CC THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING
CC GLYCOPROTEINS.
CC
CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
CC synthetic substrates.
CC
CC -1- SUBCELLULAR LOCATION: EXTERNAL, ANCHORED TO THE ENVELOPE BY ITS
CC N-TERMINAL HYDROPHOBIC SEQUENCE.
CC
CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES HEMAGGLUTININ-
CC NEURAMINIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: S58435; AAB26146.1; -
DR PIR: F48556; F48556.
DR InterPro: IPR000665; Hem-neuramidase.
DR Pfam: PF00423; HN; 1.
KW Hydroxylase; Hemagglutinin; Envelope protein; Glycoprotein;
KW Transmembrane.
FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 5 58 MEMBRANE ANCHOR (POTENTIAL).
FT DOMAIN 59 617 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 617 AA; 69322 MW; F7D48744436C1CA CRC64;
Query Match 42.68; Score 43; DB 1; Length 617;
Best Local Similarity 50.08; Pred. No. 33;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
OY 5 VMGDAESFGAIAHGGL 20
DB 584 VLADSESGHITHSGM 599
RESULT 8
HEMA_MEASE STANDARD; PRT; 617 AA.
ID HEMA_MEASE
AC P08362;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin-neuraminidase (EC 3.2.1.18).
GN HN.
OS Measles virus (strain Edmonston) (Subacute sclerosing panencephalitis
OS virus).
CC Viruses; ssRNA negative-strand viruses; Mononegavirales;
CC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
OX NCBI_TaxID=11235;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=86181590; PubMed=3008420;
RA Alkhatib G., Briedis D.J.;
RT "The predicted primary structure of the measles virus hemagglutinin.";
RT Virology 150:479-490(1986).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=90085790; PubMed=2596022;
RX Cattaneo R., Schmidt A., Spielhofer P., Kaelin K., Bacsko K.,
RA Meulen V., Pardowitz J., Flanagan S., Rima B.K., Udem S.A.;
RT "Mutated and hypermutated genes of persistent measles viruses which
RT caused lethal human brain diseases.";
RT Virology 173:415-425(1989).
CC
CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS
CC TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE
CC ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING
CC THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING
CC GLYCOPROTEINS.
CC
CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in
CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
CC synthetic substrates.
CC
CC -1- SUBCELLULAR LOCATION: EXTERNAL, ANCHORED TO THE ENVELOPE BY ITS
CC N-TERMINAL HYDROPHOBIC SEQUENCE.
CC
CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES HEMAGGLUTININ-
CC NEURAMINIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M14877; AAA46424.1; -
DR PIR: K01711; AAA75500.1; -
DR InterPro: IPR000665; Hem-neuramidase.
DR Pfam: PF00423; HN; 1.
KW Hydroxylase; Hemagglutinin; Envelope protein; Glycoprotein;
KW Transmembrane.
FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 5 58 MEMBRANE ANCHOR (POTENTIAL).
FT DOMAIN 59 617 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 617 AA; 69249 MW; 0E5A05AEDA43D9C6 CRC64;
Query Match 42.68; Score 43; DB 1; Length 617;
Best Local Similarity 50.08; Pred. No. 33;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
OY 5 VMGDAESFGAIAHGGL 20
DB 584 VLADSESGHITHSGM 599
RESULT 9
HEMA_MEASH STANDARD; PRT; 617 AA.
ID HEMA_MEASH
AC P06830;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin-neuraminidase (EC 3.2.1.18).
GN HN.
OS Measles virus (strain Halle) (Subacute sclerosing panencephalitis
OS virus).
CC Viruses; ssRNA negative-strand viruses; Mononegavirales;
CC Paramyxoviridae; Paramyxovirinae; Morbillivirus.

CC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
 OX NCBI_TaxID=11236;
 [1]
 RA SEQUENCE FROM N.A.
 RX MEDLINE=87085487; PubMed=3794664;
 RA Gerald C., Buckland R., Barker R., Freeman G., Wild T.F.;
 RT "Measles virus hemagglutinin gene: cloning, complete nucleotide
 sequence analysis and expression in COS cells.";
 RL J. Gen. Virol. 67:2695-2703(1986).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS
 TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE
 ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING
 THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING
 GLYCOPROTEINS.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
 CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in
 CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
 CC synthetic substrates.
 CC -1- SUBCELLULAR LOCATION: EXTERNAL, ANCHORED TO THE ENVELOPE BY ITS
 CC N-TERMINAL HYDROPHOBIC SEQUENCE.
 CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES HEMAGGLUTININ-
 CC NEURAMINIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X04720; CAA28427.1; -
 DR PIR: A27007; HMNZHA.
 DR InterPro: IPR000665; Hem-neuramndse.
 DR Pfam: PF00423; HN; 1.
 KW Hydroxylase; Hemagglutinin; Envelope protein; Glycoprotein;
 KW Transmembrane.
 FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 35 58 MEMBRANE ANCHOR (POTENTIAL).
 FT DOMAIN 59 617 EXTRACELLULAR (POTENTIAL).
 FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 617 AA; 69077 MW; 39240D495A525C72 CRC64;
 Query Match 42.6%; Score 43; DB 1; Length 617;
 Best Local Similarity 50.0%; Pred. No. 33;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 5 VMGDAESFGAIAHGCL 20
 DB 584 VLADSESGHITHSGM 599
 RESULT 10
 HEMA_MEASI STANDARD; PRT; 617 AA.
 ID HEMA_MEASI
 AC P26028;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hemagglutinin-neuraminidase (EC 3.2.1.18).
 GN HN.
 OS Measles virus (strain IP-3-Ce) (Subacute sclerosing panencephalitis
 OS virus).
 CC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 CC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
 OX NCBI_TaxID=11237;
 OX [1]
 RN SEQUENCE FROM N.A.

RX MEDLINE=92263801; PubMed=1585658;
 RA Schmid A., Spielhofer P., Cattaneo R., Bacsko K., Ter Meulen V.,
 RA Billeter M.A.;
 RT "Subacute sclerosing panencephalitis is typically characterized by
 RT alterations in the fusion protein cytoplasmic domain of the
 RT persisting measles virus.";
 RL Virology 188:910-915(1992).
 CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS
 TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE
 ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING
 THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING
 GLYCOPROTEINS.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of alpha-(2->3)-, alpha-(2->6)-,
 CC alpha-(2->8)-glycosidic linkages of terminal sialic residues in
 CC oligosaccharides, glycoproteins, glycolipids, colominic acid and
 CC synthetic substrates.
 CC -1- SUBCELLULAR LOCATION: EXTERNAL, ANCHORED TO THE ENVELOPE BY ITS
 CC N-TERMINAL HYDROPHOBIC SEQUENCE.
 CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES HEMAGGLUTININ-
 CC NEURAMINIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X16566; CAA34569.1; -
 DR InterPro: IPR000665; Hem-neuramndse.
 DR Pfam: PF00423; HN; 1.
 KW Hydroxylase; Hemagglutinin; Envelope protein; Glycoprotein;
 KW Transmembrane.
 FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 35 58 MEMBRANE ANCHOR (POTENTIAL).
 FT DOMAIN 59 617 EXTRACELLULAR (POTENTIAL).
 FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 215 215 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 617 AA; 66607 MW; 8F518CE44ACB4F3 CRC64;
 Query Match 42.6%; Score 43; DB 1; Length 617;
 Best Local Similarity 50.0%; Pred. No. 33;
 Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 QY 5 VMGDAESFGAIAHGCL 20
 DB 584 VLADSESGHITHSGM 599
 RESULT 11
 O2FL_HUMAN STANDARD; PRT; 317 AA.
 ID O2FL_HUMAN
 AC O13607; O9UNDX1.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Olfactory receptor 2P1 (Olfactory receptor-like protein OLF3).
 GN OR2F1 OR OLF3.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 OX [1]
 RN SEQUENCE FROM N.A.
 RA Issel-Tarver L., Rine J.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.

CC	RA	BAUER C Williams D.;	to the EMBL/GenBank/DBJ databases.
CC	DR	Submitted (JUN-1998) ;	
CC	CC	-1- FUNCTION: PUTATIVE ODORANT RECEPTOR.	
CC	CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.	
CC	CC	-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.	
CC	CC	-----	
CC	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	CC	use by non-profit institutions as long as its content is in no way	
CC	CC	modified and this statement is not removed. Usage by and for commercial	
CC	CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	CC	or send an email to license@isb-sib.ch).	
CC	CC	-----	
DR	DR	EMBL; U56421; AAB01215.1; -	
DR	DR	EMBL; AC004853; AAC64376.1; -	
DR	DR	GCRdb; GCR_1924; -	
DR	DR	InterPro; IPR000276; GPCR_Rhodpsn.	
DR	DR	Pfam; PF000001; 7tm.1; 1.	
DR	DR	PRINTS; PR00237; GPCRHHODPSN.	
DR	DR	PROSITE; PS00237; G-PROTEIN_RECEP_F1_1; 1.	
DR	DR	PROSITE; PS50262; G-PROTEIN_RECEP_F1_2; 1.	
KW	KW	G-protein coupled receptor; Transmembrane; Glycoprotein;	
KW	KW	Multigene family; Olfaction.	
FT	FT	DOMAIN 1 24	EXTRACELLULAR (POTENTIAL).
FT	FT	TRANSMEM 25 48	1 (POTENTIAL).
FT	FT	DOMAIN 49 57	CYTOPLASMIC (POTENTIAL).
FT	FT	TRANSMEM 58 79	2 (POTENTIAL).
FT	FT	DOMAIN 80 100	EXTRACELLULAR (POTENTIAL).
FT	FT	TRANSMEM 101 120	3 (POTENTIAL).
FT	FT	DOMAIN 121 139	CYTOPLASMIC (POTENTIAL).
FT	FT	TRANSMEM 140 160	4 (POTENTIAL).
FT	FT	DOMAIN 161 200	EXTRACELLULAR (POTENTIAL).
FT	FT	TRANSMEM 201 222	5 (POTENTIAL).
FT	FT	DOMAIN 223 236	CYTOPLASMIC (POTENTIAL).
FT	FT	TRANSMEM 237 261	6 (POTENTIAL).
FT	FT	DOMAIN 262 272	EXTRACELLULAR (POTENTIAL).
FT	FT	TRANSMEM 273 292	7 (POTENTIAL).
FT	FT	DOMAIN 293 317	CYTOPLASMIC (POTENTIAL).
FT	FT	DISULFID 97 189	BY SIMILARITY.
FT	FT	CARBOHYD 5 5	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	FT	CONFLICT 211 211	F -> L (IN REF. 1).
SO	SO	SEQUENCE 317 AA; 35350 MW; A72ADELAE8680CE CXC64;	
QY	QY	4 VVMGDAESFGAIGHGL 20	
DB	DB	124 VAVCDALRYSAIMHGL 140	
RESULT 12			
Y830_METTH			
ID	ID	Y830_METTH STANDARD: PRT; 329 AA.	
AC	AC	O26918;	
DT	DT	15-JUL-1998 (Rel. 36, Created)	
DT	DT	15-JUL-1998 (Rel. 36, Last sequence update)	
DT	DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	DE	Hypothetical protein MTH830.	
GN	GN	MTH830.	
OS	OS	Methanobacterium thermoautotrophicum.	
OC	OC	Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;	
OC	OC	Methanobacter.	
OX	OX	NCBI_TaxID=145262;	
RP	RP	SEQUENCE FROM N.A.	
RC	RC	STRAIN=DELTA H;	
RA	RA	MEDLINE=98037514; PubMed=9371463;	
RA	RA	Smith D.R., Doucette-Stamm L.A., Deloufery C., Lee H.-M., Dubois J.,	
RA	RA	Aldehyde T., Bashirizadeh R., Blakey D., Cook R., Gilbert K.,	

```

RA Harrison D., Hoang L., Keagle P., Lumm W., Pochier B., Qiu D.,
RA Spadatoro R., Vitarone R., Wang Y., Wierzbowski J., Gibson R.,
RA Utami N., Carruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDonough S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RL deltaH: functional analysis and comparative genomics." ;
J. Bacteriol. 179:7135-7155(1997).
CC -I- SIMILARITY: TO M.JANNASCHIT MJ1427 AND A.FULGIDUS AF2089.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000860; AAB85328.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 329 AA; 35602 MW; 30BA607BD16A5213 CRC64;
-----
OY 2 PAVVWGDAESFG 13
Query Match 41.6%; Score 42; DB 1; Length 329;
Best Local Similarity 58.3%; Pred. NO. 26;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
-----
ID RA54_YEAST STANDARD: PRT; 898 AA.
AC P32863;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE DNA repair and recombination protein RAD54.
DE RAD54 OR YGL163C OR G1821.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92009184; PubMed=1916269;
RA Emery H.S., Schild D., Kellogg D.E., Mortimer R.K.;
RT "Sequence of RAD54, a Saccharomyces cerevisiae gene involved in
RT recombination and repair." ;
RL Gene 104:103-106(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96158061; PubMed=8585224;
RA James C.M., Indge K.J., Oliver S.G.;
RT "DNA sequence analysis of a 35 kb segment from Saccharomyces
RT cerevisiae chromosome VII reveals 19 open reading frames including
RT RAD54, ACE1/CUP2, PMR1, RCK1, AMS1 and CAL1/CDC43." ;
RL Yeast 11:1413-1419(1995).
CC -I- FUNCTION: INVOLVED IN DNA REPAIR AND MITOTIC RECOMBINATION.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- SIMILARITY: BELONGS TO THE SBF2/RAD54 HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M63232; AAA34949.1; -

```

DR EMBL: Z48618; CAA88534.1; -
 DR EMBL: Z72685; CAA96875.1; -
 DR PIR: JH0440; JH0440.
 DR SGD: S0003131; RAD54.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR000330; SNF2_N.
 DR Pfam: PF00271; Helicase_C; 1.
 DR Pfam: PF00176; SNF2_N; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR DNA repair: Nuclear protein; DNA-binding; Helicase; ATP-binding.
 KW DOMAIN 21 25 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 41 45 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT NE_BIND 335 342 ATP (POTENTIAL).
 FT SITE 455 458 DECH BOX.
 SQ SEQUENCE 898 AA; 101754 MW; 2EC0A9B7E3D66292 CRC64;

Query Match 41.6%; Score 42; DB 1; Length 898;
 Best Local Similarity 40.0%; Pred. No. 67;
 Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 APAAVMDAESFGAIAHGL 20
 DB 853 APAAVMDAESFGAIAHGL 872

RESULT 14
 APTL_CONGL STANDARD; PRT; 185 AA.
 AC 087330;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Adenine phosphoribosyltransferase (EC 2.4.2.7) (APRT).
 GN APT.
 OS Corynebacterium glutamicum (Brevibacterium flavum).
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
 OC Corynebacterium.
 OX NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 13032;
 RX MEDLINE=98361031; PubMed=9695918;
 RA Wehner L., Schaefer A., Burkowski A., Kraemer R., Mechold U.,
 RA Malke H., Puchner A., Kalinowski J.;
 RT "The role of the Corynebacterium glutamicum rel gene in (p)ppGpp
 metabolism";
 RL Microbiology 144:1853-1862(1998).
 CC -1- FUNCTION: CATALYZES A SALVAGE REACTION RESULTING IN THE FORMATION
 CC OF AMP, THAT IS ENERGETICALLY LESS COSTLY THAN DE NOVO SYNTHESIS.
 CC -1- CATALYTIC ACTIVITY: AMP + diphosphate = adenine + 5-phospho-
 CC alpha-D-ribose 1-diphosphate.
 CC -1- PATHWAY: PURINE SALVAGE.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE PURINE/PYRIMIDINE
 CC PHOSPHORIBOSYLTRANSFERASE FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: AF038651; AAC35493.1; -
 DR InterPro: IPR000836; PriBosyltran.
 DR InterPro: IPR002375; Pur-pyr-pr-transf.
 DR Pfam: PF00156; PriBosyltran; 1.

DR PROSITE: PS00103; PUR_PYR_PR_TRANSFER; 1.
 KW Transferase; Glycosyltransferase; Purine salvage.
 SQ SEQUENCE 185 AA; 19563 MW; FDBAC03A3FF18182 CRC64;

Query Match 40.6%; Score 41; DB 1; Length 185;
 Best Local Similarity 72.7%; Pred. No. 21;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 VMDAESFGAI 15
 DB 39 VMDAESFGAV 49

RESULT 15
 YHQ_ECOLI STANDARD; PRT; 221 AA.
 ID YHQ_ECOLI
 AC P37619;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein yhq.
 GN YHQ OR B3471.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=94316500; PubMed=8041620;
 RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
 RT region from 76.0 to 81.5 minutes";
 RL Nucleic Acids Res. 22:2576-2586(1994).
 CC -1- SIMILARITY: STRONG, TO H-INFLUENZA H10862.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: U00039; AB018446.1; -
 DR EMBL: AE000423; AAC76496.1; -
 DR Ecogene: EG12217; yhq.
 DR InterPro: IPR003744; DUF165.
 DR Pfam: PF02592; DUF165; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 221 AA; 25269 MW; E3D90237BEF7ABCC CRC64;

Query Match 40.6%; Score 41; DB 1; Length 221;
 Best Local Similarity 53.3%; Pred. No. 25;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 6 MDAESFGAIAHGL 20
 DB 92 MGSWGFALAHFNL 106

Search completed: July 8, 2002, 08:21:24
 Job time: 184 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 8, 2002, 08:21:03 ; Search time 41.68 Seconds
(without alignments)
83.011 Million cell updates/sec

Title: US-09-582-296-2
Perfect score: 101
Sequence: 1 APAYVMGDAESFGAIHAGGL 20

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.potent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriaph:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	47.5	749	16 P73218	P73218 synechocyst
2	48	47.5	903	2 Q91255	Q91255 streptomyc
3	48	47.5	936	2 Q9AMT5	Q9AMT5 bradyrhizob
4	47	46.5	474	12 Q89542	Q89542 bovine herp
5	46	45.5	105	2 Q929R6	Q929R6 bacillus ha
6	46	45.5	121	16 Q9KE98	Q9KE98 bacillus ha
7	46	45.5	273	16 Q98106	Q98106 rhizobium l
8	45.5	45.0	899	5 Q9VWQ0	Q9VWQ0 drosophila
9	45.5	45.0	1513	16 P96901	P96901 mycobacteri
10	45	44.6	583	16 Q9PHM7	Q9PHM7 campylobact
11	45	44.6	746	17 Q9HPT7	Q9HPT7 halobacteri
12	44.5	44.1	178	6 Q46426	Q46426 oryctolagus
13	44.5	44.1	302	11 Q923U5	Q923U5 dipodomys m
14	44.5	44.1	324	11 Q923U4	Q923U4 dipodomys m
15	44.5	44.1	469	16 Q9J215	Q9J215 neisseria m
16	44.5	44.1	469	16 Q9J082	Q9J082 neisseria m

ALIGNMENTS

17	44.5	44.1	474	16 Q9PAX0	Q9PAX0 xylella fas
18	44.5	44.1	474	16 Q9HZA3	Q9HZA3 pseudomonas
19	44	43.6	610	12 Q9WJ08	Q9WJ08 measles vir
20	43.5	43.1	448	2 Q9EV53	Q9EV53 rhizobium m
21	43.5	43.1	469	16 Q98EP1	Q98EP1 rhizobium l
22	43.5	43.1	469	16 Q92L76	Q92L76 rhizobium m
23	43.5	43.1	479	16 Q9ABN0	Q9ABN0 caulobacter
24	43.5	43.1	552	10 P93837	P93837 brasica na
25	43.5	43.1	2946	10 Q64634	Q64634 arabidopsis
26	43	42.6	187	5 Q45045	Q45045 scitropnaga
27	43	42.6	244	16 Q9YJ05	Q9YJ05 streptococ
28	43	42.6	266	16 Q92W38	Q92W38 rhizobium m
29	43	42.6	407	12 Q98376	Q98376 measles vir
30	43	42.6	407	12 Q98378	Q98378 measles vir
31	43	42.6	407	12 Q98379	Q98379 measles vir
32	43	42.6	407	12 Q98380	Q98380 measles vir
33	43	42.6	407	12 Q98381	Q98381 measles vir
34	43	42.6	407	12 Q98382	Q98382 measles vir
35	43	42.6	407	12 Q98383	Q98383 measles vir
36	43	42.6	407	12 Q98384	Q98384 measles vir
37	43	42.6	407	12 Q98385	Q98385 measles vir
38	43	42.6	407	12 Q98386	Q98386 measles vir
39	43	42.6	407	12 Q98387	Q98387 measles vir
40	43	42.6	407	12 Q98388	Q98388 measles vir
41	43	42.6	407	12 Q98389	Q98389 measles vir
42	43	42.6	407	12 Q98390	Q98390 measles vir
43	43	42.6	407	12 Q98392	Q98392 measles vir
44	43	42.6	407	12 Q98393	Q98393 measles vir
45	43	42.6	407	12 Q98394	Q98394 measles vir

RESULT 1					
ID	P73218	PRELIMINARY:	PRT:	749 AA.	
AC	P73218:				
DT	01-FEB-1997 (TREMBLrel. 02, Created)				
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)				
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE	PENICILLIN-BINDING PROTEIN 1B.				
GN	MRCB OR SLR1710.				
OS	Synechocystis sp. (strain PCC 6803).				
OC	Bacteria; Cyanobacteria; Chroococcales; Synechocystis.				
OX	NCBI_TaxID=1148;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97061201; PubMed=8905231;				
RA	Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,				
RA	Miyajima N., Hirosewa M., Sugiyura M., Sasamoto S., Kimura T.,				
RA	Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,				
RA	Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,				
RA	Tabata S.;				
RT	*Sequence analysis of the genome of the unicellular cyanobacterium				
RT	Synechocystis sp. strain PCC6803. II. Sequence determination of the				
RT	entire genome and assignment of potential protein-coding regions.;				
RL	DNA Res. 3:109-136(1996).				
RL	EMBL: D90904; BAA17245.1; -				
DR	InterPro: IPR00253; FHA_domain.				
DR	InterPro: IPR001264; Transglycosyl.				
DR	InterPro: IPR001460; Transpeptidase.				
DR	Pfam: PF00498; FHA; 1.				
DR	Pfam: PF00912; Transglycosyl; 1.				
DR	Pfam: PF00905; Transpeptidase; 1.				
DR	ProDom: PD001895; Transglycosyl; 1.				
DR	SMART: SM00240; FHA; 1.				
DR	PROSITE: PS50006; FHA_DOMAIN; 1.				
KW	Complete proteome.				
SQ	SEQUENCE 749 AA; 82251 MW; 590522526832E06 CRC64;				

Query Match 47.5%; Score 48; DB 16; Length 749;

DR EMBL: AJ004801; CAA06095.1; -
SQ SEQUENCE 474 AA; 49975 MW; 63C977A0BADE6AC7 CRC64;

Query Match 46.5%; Score 47; DB 12; Length 474;
Best Local Similarity 60.0%; Pred. No. 37;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 6 MGDAESFGAIAHGCL 20
:|||||:|
DB 173 LGDASSYGAVPGGCL 187

RESULT 5
O929R6 PRELIMINARY; PRT; 105 AA.
AC O929R6; 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
DE 01-MAY-2001 (TREMBLrel. 16, last annotation update)
DE UNKNOWN (FRAGMENT).
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125;
RX MEDLINE=99184645; PubMed=10086841;
RA Takami H., Nakasone K., Hirama C., Takaki Y., Masui N., Fujii F.,
RA Nakamura Y., Inoue A.;
RT "An improved physical and genetic map of the genome of alkaliphilic
RT Bacillus sp. C-125";
RL Extremophiles 3:21-28(1999).
DR EMBL: AB013566; BAA73566.1; -.
FT NON_TER 105
SQ SEQUENCE 105 AA; 11853 MW; 7E18ACB7566A2F88 CRC64;

Query Match 45.5%; Score 46; DB 2; Length 105;
Best Local Similarity 56.2%; Pred. No. 9.5;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 5 VMGDAESFGAIAHGCL 20
:|||||:|
DB 39 MGDLEVFSHIVHAGL 54

RESULT 6
O9KE98 PRELIMINARY; PRT; 121 AA.
AC O9KE98; 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, last annotation update)
DE BH0958 PROTEIN.
GN BH0958.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fujii F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001510; BAB04677.1; -.
KW Complete proteome.

SQ SEQUENCE 121 AA; 13594 MW; ABE59AF3911A4F7B CRC64;

Query Match 45.5%; Score 46; DB 16; Length 121;
Best Local Similarity 56.2%; Pred. No. 11;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 5 VMGDAESFGAIAHGCL 20
:|||||:|
DB 39 MGDLEVFSHIVHAGL 54

RESULT 7
O98L06 PRELIMINARY; PRT; 273 AA.
AC O98L06; 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, last sequence update)
DE 01-OCT-2001 (TREMBLrel. 18, last annotation update)
DE MLR1236 PROTEIN.
GN MLR1236.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP002996; BAB48657.1; -.
DR InterPro: IPR002376; formyl_transf.
DR Pfam: PF00551; formyl_transf. 1.
KW Complete proteome.
SQ SEQUENCE 273 AA; 29214 MW; 6975846D61F5E2D2 CRC64;

Query Match 45.5%; Score 46; DB 16; Length 273;
Best Local Similarity 53.3%; Pred. No. 28;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 AVVMGDAESFGAIAH 17
:|:|||||:|
DB 177 ALVSGDAQNFQTVH 191

RESULT 8
O9VNO0 PRELIMINARY; PRT; 899 AA.
AC O9VNO0; 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE CG10712 PROTEIN.
GN CG10712 OR CG18622 OR CG18623.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

Query Match:	45.0%;	Score 45.5;	DB 5;	Length 899;
Best Local Similarity	55.0%;	Pred. No. 1.3e+02;		
Matches 11;	Conservative	2;	Mismatches 6;	Indels 1;
				Gaps 1
QY	2 PAVVMGDA-ESFGAIAHGL 20	: :		
		: :		
Db	82 PWIVAGKTIIQSFQIAHGQL 101			
RESULT	9			
ID	P96901	PRELIMINARY;	PRT;	1513 AA.
AC	P96901;			
DT	01-MAY-1997 (TREMBLrel. 03, Created)			
DT	01-MAY-1997 (TREMBLrel. 03, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	COSMID Y71.			
GN	RV3296 OR MTCY71.36.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteriiae;			
OC	Actinomycetales; Corynebacteriineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_Taxid=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RX	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			

Query Match:	45.0%	Score 45.5:	DB 16:	Length 1513:
Best Local Similarity	48.8%	Pred. No.2,4e+02:		
Matches 11;	Conservative 3;	Mismatches 5;	Indels 5;	Gaps 1;
Qy	1	APAAVWGDAEFG-----A	AHAGG	19
Db	316	APAHIMSGQTGCAPYLARAHHG		339

RESULT	10			
09PHM7				
ID	09PHM7	PRELIMINARY:	PRF:	583 AA.
AC	09PHM7			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	ASPARTYL-TRNA SYNTHETASE (EC 6.1.1.12).			
GN	ASPS OR CU0640.			
OS	Campylobacter jejuni.			
OC	Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;			
OC	Campylobacter.			
OX	NCBI_TaxID=197;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN=NCTC 11168;			
RX	MEDLINE=20150912; PubMed=10688204;			
RA	Barthill J., Wen B.W., Mungall K., Ketley J.M., Churcher C.,			
RA	Pasham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,			
RA	Quail M.A., Karyshev A.V., Mole S., Pallen M.J., Penn C.W.,			
RA	Whitehead S., Barrett B.G.;			
RT	"The genome sequence of the food-borne pathogen Campylobacter jejuni			
RT	reveals hypervariable sequences."			
RL	Nature 403:665-668(2000).			
CC	-1- CATALYTIC ACTIVITY: ATP + L-AMINO ACID + TRNA(AMINO ACID) = AMP +			
CC	PYROPHOSPHATE + L-AMINOACYL-TRNA(AMINO ACID).			
CC	-1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE (OR L-ASPARAGINE) +			
CC	TRNA(ASN) = AMP + PYROPHOSPHATE + L-ASPARTYL-TRNA(ASP) (OR L-			
CC	ASPARAGINYL-TRNA(ASN)).			
CC	-1- SUBUNIT: HOMODIMER (BY SIMILARITY).			
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.			
CC	-1- SIMILARITY: BELONGS TO ASPARTYL-TRNA SYNTHETASE FAMILY.			
DR	EMBL: AL13075; CAB75276.1; -.			
DR	HSSP: P21889; 1EOR.			
DR	InterPro: IPR002106; AA_trna_ligase_II.			
DR	InterPro: IPR002312; trna-synt_asp			

DR InterPro: IPR003313; tRNA-synt_Lys_2.
 DR InterPro: IPR004365; tRNA_antl.
 DR Pfam: PF01336; tRNA_antl. 1.
 DR PRINTS: PR01042; TRNASYNTHASP.
 DR PRINTS: PR00982; TRNASYNTHASP.
 DR PROSITE: PS00179; AA_TRNA_LIGASE_II_1; 1.
 DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN.1.
 DR ATP-binding; Aminoacyl-tRNA synthetase; Complete proteome; Ligase;
 KW Protein biosynthesis.
 SQ SEQUENCE 583 AA; 66166 MW; A3723A2E100222E2 CRC64;

Query Match 44.6%; Score 45; DB 16; Length 583;
 Best Local Similarity 69.2%; Pred. No. 96;
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 8 DAESFGAIAHGGL 20
 |||||
 DB 514 DALSGAPPHGCI 526

RESULT 11
 O9HPT7 PRELIMINARY; PRT; 746 AA.
 AC O9HPT7;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE VNG1476C.
 GN VNG1476C.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE:20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahaira G.G., Bergquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shrogha J.,
 RA Swartzell S., Weir D., Hall T.A., Weller R., Goo Y.A.,
 RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angwine C.M., Dale H.,
 RA Isehnager T.A., Peck R.F., Pohlshroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
 RT "Genome sequence of Halobacterium species NRC-1";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL: AE005063; AAC19780.1;
 DR InterPro: IPR000731; HMGCR_patched_57M.
 DR PROSITE: PS50156; SSD; 2.
 KW Complete proteome.
 SQ SEQUENCE 746 AA; 79248 MW; F278AC41A19B9555 CRC64;

Query Match 44.6%; Score 45; DB 17; Length 746;
 Best Local Similarity 50.0%; Pred. No. 13e+02;
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 3 AAVMGDAESFGAIAHGGL 20
 |||||
 DB 220 AVEVGVIYTFGAIGYCGI 237

RESULT 12
 O46426 PRELIMINARY; PRT; 178 AA.
 AC O46426;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE AOUAPORIN 4 (FRAGMENT).
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NEW ZEALAND;
 RA Carter E.P., Umenishi F., Matthey M.A., Verkmann A.S.;
 RT "Increased water permeability across the blood-gas barrier in rabbit
 lungs in the first 24 hours after birth."
 RL J. Clin. Invest. 0:0-0(1997).
 DR EMBL: AF000312; AB94409.1;
 DR HSSP: P29972; 1FOY.
 DR InterPro: IPR000425; MIP.
 DR Pfam: PF00230; MIP; 1.
 DR PRINTS: PR00783; MINTRNLSICP.
 KW Porin.
 FT NON_TER 1
 FT NON_TER 178
 SQ SEQUENCE 178 AA; 19236 MW; CE43FE18ACBAE2 CRC64;

Query Match 44.1%; Score 44.5; DB 6; Length 178;
 Best Local Similarity 39.3%; Pred. No. 30;
 Matches 11; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

OY 2 PAVVMGDAES-----FGAIAHGGL 20
 |||||
 DB 82 PAVIMGNMNMHWIYWGPIIGAVIAGGL 109

RESULT 13
 O92335 PRELIMINARY; PRT; 302 AA.
 AC O92335;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE AOUAPORIN 4 ISOFORM 1.
 OS Dipodomys merriami (Merriam's kangaroo rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Heteromyidae;
 OC Dipodomysinae; Dipodomys.
 OX NCBI_TaxID=94247;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fang P.K., Huang Y., Walsberg G.E., Brown D., van Hoek A.N.;
 RT "Molecular Cloning and Characterization of Merriam's Kangaroo Rat
 Aquaporin 4";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY032857; AAK66823.1;
 KW Porin.
 SQ SEQUENCE 302 AA; 32439 MW; EECDB5A25799CD2 CRC64;

Query Match 44.1%; Score 44.5; DB 11; Length 302;
 Best Local Similarity 39.3%; Pred. No. 55;
 Matches 11; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

OY 2 PAVVMGDAES-----FGAIAHGGL 20
 |||||
 DB 198 PAVIMGNMNMHWIYWGPIIGAVIAGGL 225

RESULT 14
 O92334 PRELIMINARY; PRT; 324 AA.
 AC O92334;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE AOUAPORIN 4 ISOFORM 2.
 OS Dipodomys merriami (Merriam's kangaroo rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Heteromyidae;

OC Dipodomysinae; Dipodomys.
 OX NCBI_TaxID=94247;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Pang P.K., Huang Y., Walsberg G.E., Brown D., van Hoek A.N.;
 RT "Molecular Cloning and Characterization of Merriam's Kangaroo Rat
 Aquaporin 4.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY032858; AAK66824.1; -
 KW Porin.
 SQ SEQUENCE 324 AA; 34939 MW; EAE9CE3B17952CDD CRC64;

Query Match 44.1%; Score 44.5; DB 11; Length 324;
 Best Local Similarity 39.3%; Pred. No. 59;
 Matches 11; Conservative 4; Mismatches 4; Indels 9; Gaps 1;

OY 2 PAVMGDAES-----FGAIAHGCL 20
 DB 220 PAVMGWENHWIYWGPIIGAVLAGCL 247

RESULT 15

O9JZ15 PRELIMINARY; PRT; 469 AA.
 AC O9JZ15;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 3-ISOPROPYLMALATE DEHYDRATASE, LARGE SUBUNIT.
 GN NMB1036.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / SEROGROUP B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciltone H., Clark E.B.,
 RA Cotton M.D., Ulteback T.R., Khouri H., Qin H., Yamathavan J.,
 RA Gill J., Scariato V., Maignani V., Pizze M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 MC58.";
 RL Science 287:1809-1815(2000).
 DR EMBL; AE002454; AAF41435.1; -.
 DR TIGR; NMB1036; -.
 DR InterPro; IPR001030; Aconitase.
 DR Pfam; PF00330; aconitase; 1.
 DR PRINTS; PR00415; ACONITASE.
 DR ProDom; PD000511; ACONITASE.
 DR PROSITE; PS00450; ACONITASE_1; 1.
 DR PROSITE; PS01244; ACONITASE_2; 1.
 KW Complete proteome.
 SQ SEQUENCE 469 AA; 50843 MW; D538C334E590E4E0 CRC64;

Query Match 44.1%; Score 44.5; DB 16; Length 469;
 Best Local Similarity 50.0%; Pred. No. 90;
 Matches 10; Conservative 3; Mismatches 2; Indels 5; Gaps 1;

OY 4 VVMGDAE-----SFGAIAHG 18
 DB 126 VVCGDSHTSTHGAFCALAHG 145

Search completed: July 8, 2002, 08:21:05
 Job time: 220 sec

Mon Jul 8 09:33:19 2002

us-09-582-296-2.rspt

